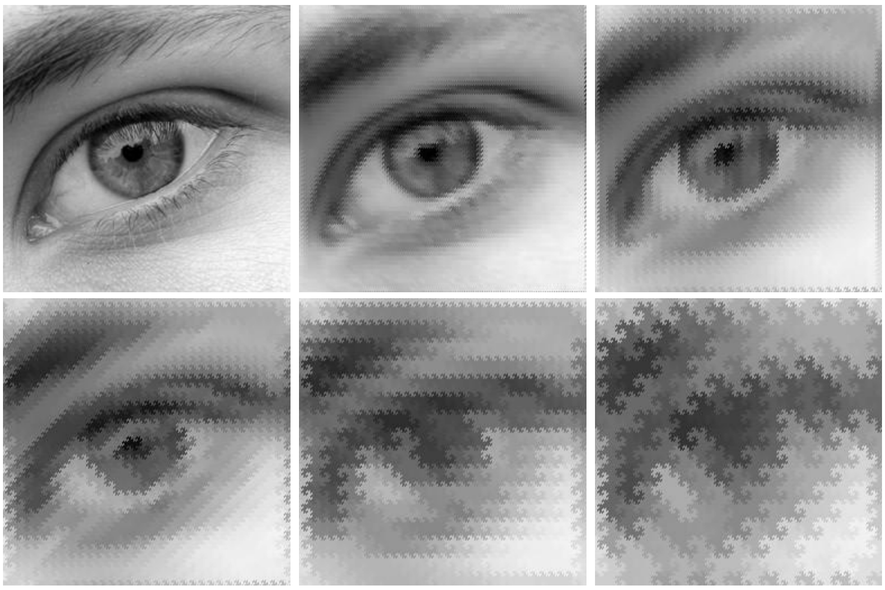
Mathias Lindemann

Approximation Properties of Non-Separable Wavelet Bases with Isotropic Scaling Matrices

and their Relation to Besov Spaces



University Bremen, 2005

APPROXIMATION PROPERTIES OF NON-SEPARABLE WAVELET BASES WITH ISOTROPIC SCALING MATRICES

AND THEIR RELATION TO BESOV SPACES

von Mathias Lindemann

Dissertation

zur Erlangung des Grades eines Doktors der Naturwissenschaften
– Dr. rer.nat. –

Vorgelegt im Fachbereich 3 (Mathematik & Informatik)
der Universität Bremen
im Februar 2005

Datum des Promotionskolloquiums: 4. April 2005

Gutachter:	Prof. Dr. Peter Maaß	(Universität Bremen)
	Prof. Dr. Stephan Dahlke	(Universität Marburg)

Table of Contents

Zusammenfassung

In dieser Dissertation werden nichtseparable Waveletbasen und ihre Verbindung zu Besov-Räumen untersucht. Die bekannte Charakterisierung von Besov-Räumen über dyadische, separable Wavelet-Darstellungen von Funktionen wird für solche Fälle erweitert, in denen die Dilation durch eine allgemeine, ganzzahlige Matrix gegeben ist, die expandierend und isotrop ist. Neben der bekannten Quincunx-Matrix und der Box-Spline-Matrix werden noch weitere Beispiele solcher Matrizen sind angegeben.

Anwendungen in der Bildverarbeitung zeigen, daß separable Waveletbasen sehr effizient und schnell zur Analyse von Signalen unterschiedlichster Art geeignet sind. Anwendungen, die richtungsunabhängige Strukturerkennung erfordern, zum Beispiel in der Texturerkennung, bestätigen, daß separable Waveletbasen nur im begrenzten Maße ihren Dienst dafür tun. Erst nichtseparable Wavletbasen sind in der Lage, Strukturen, die nicht nur horizontal, vertikal oder diagonal sind, fein genug aufzulösen und zu erkennen.

Von mehreren Forschergruppen wurden verschiedene Ansätze zur Konstruktion von Wavelets mit allgemeinen Skalierungsmatrizen und mit separablen Waveletbasen verglichen.

Unbekannt war bisher, wie sich Beweise zur Approximationstheorie für den dyadischen, separablen Fall auf den generellen, im allgemeinen nichtseparablen Fall mit beliebigen Skalierungsmatrizen übertragen lassen. In dieser Arbeit wird die Lücke zwischen separabler und nichtseparabler Waveletapproximation geschlossen. Es wird gezeigt, welche Besov-Räume über nichtseparable Waveletbasen charakterisiert werden können.

Diese Arbeit besteht aus zwei Hauptteilen: lineare und nichtlineare Approximation mit Waveletbasen. Zuerst werden Approximations- und Glattheitseigenschaften für shiftinvariante Räume nichtseparabler Skalierungsfunktionen untersucht. Im Einzelnen werden geeignete Jackson- und

BERNSTEIN-Ungleichungen für diese Räume bewiesen. Sie sind die wichtigsten Bestandteile der allgemeinen Theorie zur Charakterisierung von BESOV-Räumen über Waveletentwicklungen. Am Ende des ersten Teils wird eine Normäquivalenz zwischen einer diskreten BESOV-Norm und einer gewichteten Norm von Waveletkoeffizienten angegeben.

Im zweiten Teil werden die Approximationsräume zur N-Term Approximation mit nichtseparablen Waveletbasen untersucht. Dabei dienen die Normäquivalenzen aus dem ersten Teil dazu, eine adaptive Auswahl von N Waveletkoeffizienten zu treffen und die zugehörige Approximationsrate anzugeben. Es wird gezeigt, daß Räume von Funktionen mit derselben Approximationsrate wieder BESOV-Räume sind.

Abstract

In this thesis we investigate the connection between non-separable wavelet bases and Besov spaces. The well known results about the characterization of Besov spaces via dyadic wavelet expansions are extended for those cases where the dilation is given by a general expanding isotropic integer matrix. Beside the Quincunx matrix or the Box-spline matrix we present other scaling matrices for non-separable wavelets.

Applications in image processing show that separable wavelet expansions are very useful for efficient and fast algorithms for the analysis of various signals. Nevertheless, particular applications such as texture recognition exhibit that separable wavelet expansions have some shortcomings. For these purposes non-separable wavelets are capable to detect sufficiently precise structures that are not only horizontal, vertical or diagonal but arbitrarily orientated.

Diverse approaches to construct non-separable wavelets had been successfully developed by other groups and were already compared to separable wavelet bases.

So far it is not known how the proofs of the approximation theory can be adopted from the dyadic separable case to the more general non-separable case with arbitrary scaling matrices. In this thesis we close the gap between separable and non-separable wavelet approximation. We show which Besov spaces can be characterized by non-separable wavelet expansions.

This thesis consists of two main parts: linear and nonlinear wavelet approximation. First we investigate approximation and smoothness properties of shift-invariant spaces generated by non-separable scaling functions. In particular we prove suitable Jackson and Bernstein estimates for these spaces. They are the most important ingredients for the general theory for the characterization of Besov spaces via wavelet expansions. At the end of

the first part we obtain a norm equivalence between a discrete version of a Besov norm and a weighted sequence norm of wavelet coefficients.

In the second part we investigate approximation spaces for the N-term approximation with non-separable wavelet bases. Here we use the norm equivalences from the first part to present an adaptive choice of N wavelet coefficients and to determine the rate of approximation. It will be shown that spaces with the same approximation rate are again Besov spaces.

Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

I would like to thank my supervisor Prof. Peter Maaß in the first instance for the support, his invitation to develop visions and ideas as well as for his encouragement to go ahead with my work.

I am deeply indebted to my supervisor Prof. Stephan Dahlke from Universität Marburg whose continuous help, stimulating suggestions and steady encouragement helped me substantially in all the time of research for and writing of this thesis. Also I thank the whole team of his group including Torsten Raasch and Karsten Koch for their assistance during the productive time in Marburg in January 2004.

I am bound to Franklin Mendivil from Acadia University for his help to compute the images in this thesis. Special thanks go to Prof. Peter Oswald from IUB Bremen for illuminating conversations and valuable comments.

I have furthermore to thank my colleagues from the Zentrum für Technomathematik for all their help, support, interest and valuable hints. Especially I am obliged to Esther Klann, Dirk Lorenz, Lutz Justen, Silva Fischer, Gerd Teschke, Henning Thielemann and Kristian Bredies.

Last but not least I am deeply grateful to Ricarda Hoffmann, Hriska and Peter Hoffmann as well as to Fabian Bahr for their personal advices and patient sympathy which enabled me to finalize this work.

Bremen, 12th April 2005

Mathias Lindemann
Zentrum für Technomathematik
Universität Bremen

Zentrum für Technomathematik

http://www.math.uni-bremen.de/~lindeman

Chapter 1

Introduction

If you ask a person working in wavelet theory why wavelets became famous during the last fifteen years you probably get different answers.

An expert of filter theory very likely would answer that wavelets provide efficient and fast algorithms to represent a signal split in its distinct frequency bands via the idea of a multiresolution analysis. Properties of specific wavelets such as orthogonality, symmetry, short length and good attenuation allow to design perfect reconstruction filter banks for audio and video compression, echo cancellation, radar, image analysis, medical image analysis, etc. The connection between discrete wavelet and scaling coefficients and efficient filter banks was introduced e.g. in Strang and Nguyen (1997) or Vetterli and Kovačević (1995).

Time-scale followers will emphasize that the wavelet transform is optimal local both in time and scale. Small scales are analyzed with a time-scale window that is small in time and large in scale in comparison to a fixed analysis window in the Fourier kingdom, see e.g. Mallat (1998) or Holschneider (1995).

Those who apply the wavelet transform in digital signal and image processing to compression and denoising would point out the approximation properties of wavelet expansions which offer the relation between data properties and wavelets. Also for image restoration or enhancement wavelets are able to approximate even non-smooth regions in an image adaptively with a minimum of data, see DeVore, Jawerth, and Popov (1992).

Quite recently also the application of wavelets to the numerical treatment of elliptic operator equations has become more and more the center of attraction. Indeed, it has turned out that the strong analytical properties of wavelets are useful to derive powerful numerical schemes including very efficient adaptive algorithms, we refer, e.g., to Dahmen (1997), Dahmen (2001), Dahlke, Dahmen, and DeVore (1997) or Dahlke, Hochmuth, and Urban (2000) for an overview.

Also, wavelet schemes were celebrated for their success with multilevel techniques for preconditioning linear systems that arise from Galerkin methods for elliptic boundary value problems and saddle point problems to accelerate the convergence of iterative methods, see e.g. Dahmen and Kunoth (1992) and Dahlke, Dahmen, and Urban (2002) or Barinka et al. (2001).

A remarkable discovery was the use of wavelet analysis for the investigation of Besov regularity for a large class of elliptic problems, especially for second order partial differential equations. In general it is hard to prove regularity results of the unknown underlying function that solves these equations. The proofs for regularity of solutions are much more elegant and easier to establish if the function is decomposed in terms of wavelets. Then the regularity results can be shown by using the characterization of smoothness spaces such as Besov spaces via norm equivalences by certain weighted sequence norms of wavelet coefficients, see e.g., Dahlke and DeVore (1997) or Dahlke (2003), Dahlke (1999b), Dahlke (1999a).

The success of numerical wavelet methods relies basically on the following fundamental properties of wavelets:

- The vanishing moments of wavelets remove the smooth part of a function;

- For a wide class of operators their representation in the wavelet basis is nearly diagonal;

- Weighted sequence norms of wavelet expansion coefficients are equivalent in a certain range to Besov norms.

In particular we emphasize the last point. This remarkable characteristic allows to determine exactly the approximation rate of a function approximated by wavelets. In general, the rate of approximation depends on the smoothness of the underlying function in a certain scale of Besov spaces, that is approximation order $\mathcal{O}(n^{-s/d})$ in $L_p(\mathbb{R}^d)$ from linear approximation spaces with dimension n is equivalent to smoothness order s in $L_p(\mathbb{R}^d)$.

A similar result is known from finite element approximation: smoothness properties of a function imply certain approximation properties, i.e. for finite element spaces V_h defined from a regular conforming partition $\mathcal{T}_h$ of a domain $\Omega \subset \mathbb{R}^d$ with uniform mesh size h and a function $f \in L_p(\mathbb{R}^d)$ such that $D^\alpha f \in L_p(\mathbb{R}^d), |\alpha| \leq t$ one has

$$f \in W^{t+s}(L_p(\Omega)) \Longrightarrow \inf_{g \in V_h} \|f - g\|_{W^t(L_p(\Omega))} \lesssim \mathcal{O}(h^s). \tag{1.0.1}$$

In this case we say that the spaces V_h provide *approximation order s*.

A surprising fact is that wavelets can be used to describe the converse direction as well: a certain rate of approximation implies a specific order of smoothness. Indeed, here lies the great capability of wavelets. They provide an analysis tool to obtain both directions and thus a full characterization of a large number of smoothness classes, namely BESOV spaces.

In comparison to the above mentioned linear approximation order obtained by a uniform grid refinement wavelets also have great success with adaptive methods. Adaptive methods can be interpreted as a nonlinear approximation strategy. Here we do not approximate by elements from a linear space but from a nonlinear manifold. The number of free parameters n, that is the dimension of the approximation space in the linear case, is replaced by the number of elements used in the nonlinear counterpart. In this version an approximation of order s is characterized by smoothness of the underlying function of order s in $L_\tau(\mathbb{R}^d)$ with $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$. Since $\tau < p$ we expect higher order of smoothness measured in $L_\tau(\mathbb{R}^d)$ rather than in $L_p(\mathbb{R}^d)$ and thus a higher rate of approximation. Functions with singularities or jumps still have a high order of smoothness for small τ. Thus, these functions are exactly those types for which nonlinear adaptive methods perform better than linear methods.

With great success these results had been fully investigated for separable, dyadic wavelet expansions with the scaling matrix $M = 2\mathcal{I}$ (by $\mathcal{I}$ we denote the identity matrix) which are obtained by tensor products generated from univariate functions. These decompositions offer efficient and powerful algorithms for most of the above mentioned applications. Moreover, their way of functionality was completely investigated and is well understood. Nevertheless, it turned out that selected items suggest the use of non-separable wavelet decompositions due to their specific properties. For instance, in texture analysis non-separable wavelets present by their isotropy a rotation invariant analysis system. Also, computational aspects require a small number of wavelets to compute. This is achieved by using scaling matrices with a small determinant.

The main part of available literature concerned with non-separable wavelets offers construction principles, see e.g. Cohen and Daubechies (1993), de Boor,

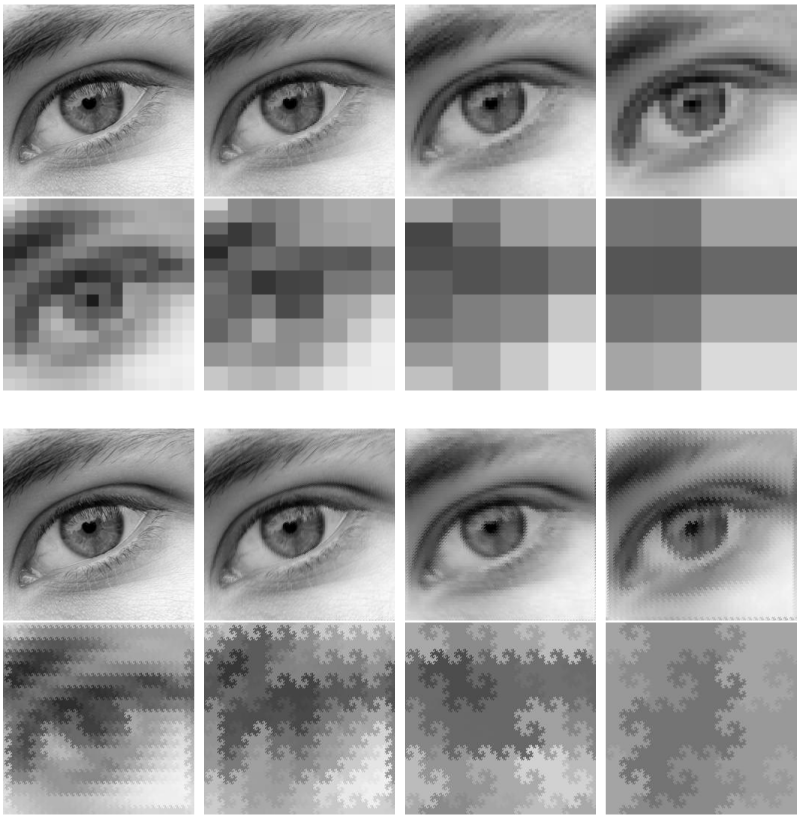


Figure 1.1: Top two rows: Separable HAAR wavelet decomposition with truncation of the top wavelet levels 1 to 7, i.e. these levels are set to zero. Bottom two rows: Non-separable QUINCUNX decomposition with truncation of the top levels 3,5,7,9,11,13,15. (Software for non-separable transform written by F. Mendivil (Acadia University, Canada)).

DeVore, and Ron (1993), Dahlke (1997), Han (2002), Han (2003), Belogay and Wang (1999), Bownik (2001) or He and Lai (2000). The construction with focus on interpolating non-separable wavelets was established in Dahlke and Maaß (1997), Dahlke, Maaß, and Teschke (2003), Derado (2001), Derado (1999), or Jia and Micchelli (1992).

Approximation and smoothness properties were investigated in Han and Jia (2002), Jia (1999) and Cohen, Gröchenig, and Villemoes (1999) or Feilner, Van De Ville, and Unser (in press). A consideration of stationary subdivision schemes and transition operators for general scaling matrices can be found in Han and Jia (1998), Jia and Zhang (1999) or Latour, Müller, and Nickel (1998). Also, a number of experimental results for the application of non-separable wavelet decompositions in image compression and texture analysis with comparison to separable wavelets are described in Andrews and Nguyen (1998), Andrews and Nguyen (2002), Van der Auwera, Munteanu, and Cornelis (2000), Lin and Ling (2003) and Mojsilovic and Popovic (1998), Mojsilovic, Markovic, and Popovic (1997), van de Wouwer (1998). A recent work where non-separable wavelets have a great success for tomographic reconstruction is Bonnet et al. (2000). The implementation of algorithms for the non-separable wavelet transform turns out to be more sophisticated due to the rotational component and the often fractional scaling factor. How to implement such a decomposition scheme for non-tensor wavelet systems is carried out in Mendivil and Piché (1999) (cf. Figure 1.1 for example). Properties of non-separable multi-dimensional perfect reconstruction filter banks are shown in Kovačević and Vetterli (1992).

In this scientific paper we close the gap between approximation properties and the connection to smoothness spaces for non-separable wavelet decompositions. For the generalization of the approximation properties for non-separable wavelet decompositions many well known concepts can be employed from the separable setting. Nonetheless there are some crucial parts, e.g. for the norm and determinant of the d-dimensional scaling matrix in the dyadic case one has $|\det M| \cdot \|M^{-1}\|_2^d = 2^d 2^{-d} = 1$ which vanishes often when rescaling arguments are applied due to its simple structure. For a general scaling matrix one only has $\|M\|_2 \cdot \|M^{-1}\|_2 \geq 1$. Similarly, terms such as $|\det M|$ and $\|M^{-1}\|_2^d$ do not cancel down for general scaling matrices and require particular consideration.

To the author's knowledge there is so far no completely elaborated reference about norm equivalences between discrete Besov spaces and non-separable wavelet expansions. This thesis is devoted to collect results about approximation theory for non-separable wavelet expansions which also include as a special case the dyadic wavelet approximation.

1.1 Organization of the Thesis

The objective of this thesis is to extend the results about norm equivalences between separable dyadic wavelet norms and discrete Besov norms to the case where the multiresolution analysis is generated by non-separable scaling functions with a scaling matrix more general than $2\mathcal{I}$. Moreover, the classical results about nonlinear approximation with separable wavelet bases are upgraded to the non-separable case.

The thesis is divided in eight chapters. After this introduction the basic definitions of non-separable scaling functions and wavelets with their corresponding multiresolution analysis as well as the difference to the dyadic version is shown in Chapter 2. First we introduce some useful notations in Section 2.2. Then we present the background story of orthogonality and stability of refinable functions in Section 2.3. After that, in Section 2.4 we prove stability results for non-separable scaling functions and wavelets. We close this Chapter with a basic introduction into smoothness spaces (Section 2.5) that play a major role in this work.

In Chapter 3 we have a detailed look on shift-invariant spaces with respect to their approximation properties. The goal of that chapter is to give an interpretation of the polynomial reproduction properties of the shift-invariant spaces generated by the scaling functions. In particular approximation properties of the level spaces V_j will be expressed by the STRANG-FIX conditions. Then we introduce an important type of inequality that is essential to build the general approximation theory: the JACKSON or *direct estimate*.

Another essential ingredient for the general approximation theory is the BERNSTEIN or inverse estimate which takes into account the smoothness properties of the shift-invariant spaces V_j. In Chapter 4 we consider BESOV spaces in a general manner. In particular we prove several versions of the BERNSTEIN estimate.

The goal of Chapter 5 is to describe the connection between approximation and smoothness properties of shift-invariant spaces. Given the wavelet coefficients of a function f we relate a weighted sequence of these coefficients to some classical notion of smoothness satisfied by f. At this point the main result is shown: the characterization of BESOV classes in terms of non-separable wavelet expansions which is the norm equivalence between a weighted sequence of wavelet coefficients and a discrete BESOV norm.

Afterwards, we employ in Chapter 6 a general duality principle to have a characterization of BESOV classes with negative smoothness: the dual wavelet basis provides wavelet coefficients for these BESOV spaces.

For the nonlinear approximation results in Chapter 8 some specific BESOV spaces turn out to be crucial. In particular for numerical wavelet schemes it is important to consider BESOV classes with smoothness order s in L_τ for $0 < \tau < 1$ which is often referred to as *unstable approximation*. The characterization of BESOV spaces by means of wavelet coefficients in this setting requires particular attention since the corresponding projectors P_j

are not necessarily bounded in L_τ for $\tau < 1$. Some techniques to circumvent these difficulties are shown in Chapter 7.

In Chapter 8 we consider N-term approximation for non-separable wavelet expansions. There we approximate a function by elements from a nonlinear manifold rather than from linear shift-invariant spaces. The free parameter is replaced by the cardinality of this manifold in comparison to the dimension of the shift-invariant spaces in the linear case. We use the norm equivalences from Chapter 5 and 6 to prove that there exists an adaptive strategy which uses a nonlinear refinement scheme. It will turn out that adaptivity pays for functions with local non-smooth parts such as jumps or singularities. As a surprising fact we will see that in dependence of the smoothness order of a function its rate of approximation by nonlinear methods is exactly described by a specific scale of BESOV spaces which turn out to be the perfect matching spaces for adaptive and nonlinear approximation methods.

Chapter 2

Multiresolution Analysis with General Scaling Matrices

The aim of this chapter is to provide the basic definitions of a multivariate non-separable multiresolution analysis. Although many aspects of the generalization from one-dimensional wavelets to the multivariate case can be worked out by tensor products there are several items that need particular attention. For example, the supports of non-separable wavelets are generally not cubes which requires proper multi-dimensional consideration. The difference between a non-separable multiresolution analysis and a tensor product type is pointed out and consolidated by examples. We do not regard construction principles of non-separable wavelets. For this the reader is referred to Dahlke (1997) or Wojtaszczyk (1997, Chap.5).

2.1 Multiresolution Analysis

The concept of multiresolution analysis is a standard tool to derive sharp approximation results. Therefore we start with an introduction into this tool. We first introduce a quite general definition. Then we study stability and orthogonality properties.

The intuitive background of the concept of multiresolution analysis comes from signal processing. It is desired to model the successive approximation of a signal with different resolutions. With the reasonable assumption of finite energy of a signal we stick to the model $f \in L_2(\mathbb{R}^d)$. Then, the different resolution levels are described by projections onto suitable subspaces of $L_2(\mathbb{R}^d)$. Since high resolution results in a better approximation we claim that these subspaces are nested. Finally, if we admit arbitrary high resolution we can reproduce the signal. That is, the union of subspaces should be dense in $L_2(\mathbb{R}^d)$. Further demands such as shifting the signal results in shifting its coefficients, etc., lead to the pyramidal algorithm for the one-dimensional multiresolution analysis which was introduced by Mallat and Daubechies, see e.g. Mallat (1998) and Daubechies (1992).

The concept of a separable multiresolution analysis carries over the properties of univariate functions to multivariate versions via the tensor product approach. Although this concept provides simple and well understood algorithms there are certain advantages for nonseparable decompositions. Namely, natural images do not necessarily exhibit horizontal or vertical structures. For instance the leaves of a tree may be oriented in any direction. Medical images, such as liver scans (see e.g. Mojsilovic and Popovic (1998)) or industrial recordings like texture images of tissues (see e.g. Mojsilovic, Markovic, and Popovic (1997)) are better handled by a proper two-dimensional transform which takes into account areas rather than rows and columns.
Andrews and Nguyen (2002, p.2) consider this matter as follows:

> *„Separable wavelet decompositions have vertical and horizontal cut-off while the non-separable decomposition can have a cut-off at an angle. This is better psychovisually because it means that the perceptually least valuable component of vision is quantized first.“*

The most common way to extend the one-dimensional multiresolution analysis to two dimensions is to apply two 1-D decompositions separately - horizontal and vertical. From a univariate scaling function φ, the multivariate version for d dimensions is obtained via tensor products by

$$\phi(x) = \varphi(x_1) \cdot \ldots \cdot \varphi(x_d) \text{ for } x = (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

If a function ϕ has such a structure we call it *separable*. Otherwise we say ϕ is *non-separable*.

Images are transformed in this way by decompositions along rows and columns via the pyramidal algorithm of Mallat (1998). This corresponds to a filtering of the image with subsampling factor 2 which is also known as a *dyadic multiresolution analysis.* As a result one obtains three spatially oriented subimages that constitute the high pass filtered versions in horizontal, vertical and diagonal direction.

The basic key for a multiresolution analysis is the generating scaling function which satisfies a so called *refinement equation.* A separable function $\phi : \mathbb{R}^d \longrightarrow \mathbb{R}$ is called *refinable* with refinement mask $(h_k)_{k\in\mathbb{Z}^d} \in \ell_2(\mathbb{Z}^d)$ if

$$\phi(x) = \sum_{k\in\mathbb{Z}^d} h_k \phi(2x-k) \qquad \text{for all } x \in \mathbb{R}^d. \tag{2.1.1}$$

A natural extension to the non-separable case is to replace the scaling factor 2 (or more precisely the scaling matrix $2\mathcal{I}$) by a general *scaling matrix.*

Then the so-called *scaling matrix* or *dilation matrix*

$$M \in \mathbb{Z}^{d\times d} \tag{2.1.2}$$

plays an important role for the properties of the corresponding multiresolution analysis. To make sure that we get a nested sequence of multiresolution spaces and that the sampling grid is mapped only onto integer points we always claim that M is an integer matrix and expanding, that is, all its eigenvalues $\lambda_1, \ldots, \lambda_d$ have modulus larger than one, or equivalently $\lim_{j\to\infty} \|M^{-j}\|_2 = 0$. The case where $M = 2\mathcal{I}$ is often referred to as a dyadic multiresolution analysis and is, by far, the most well studied case with many available results.

The notion of refineability in a general sense is expressed by the following two-scale relation.

Definition 2.1.1 (Refinable). *A non-separable scaling function ϕ is called* $(\mathbf{h}, M)$-refinable *if it satisfies a two-scale relation of the form*

$$\phi(x) = \sum_{k\in\mathbb{Z}^d} h_k \phi(Mx-k) \quad a.e. \tag{2.1.3}$$

with refinement mask $\mathbf{h} = \{h_k\}_{k\in\mathbb{Z}^d} \in \ell_2$ and the scaling matrix M.

If the mask $\mathbf{h}$ is finitely supported and satisfies

$$\sum_{k\in\mathbb{Z}^d} h_k = |\det M|, \tag{2.1.4}$$

then it is well known that there exists a unique compactly supported distribution ϕ satisfying the refinement equation (2.1.3) provided that $\hat{\phi}(0) = 1$

(see Cavaretta, Dahmen, and Micchelli (1991, Chap.5)). This distribution is called the normalized solution to the refinement equation. In the following we always assume that ϕ is normalized in this sense.

One and probably also the first approach for a non-separable wavelet decomposition is the generalization of the HAAR function.

Here the scaling matrix is given by the QUINCUNX[1] matrix $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Geometrically speaking this represents a rotation by 45° and a dilation by factor $\sqrt{2}$, see Figure 2.1 and 2.3 for example.
In this case we look for a function

$$\phi(x) = \chi_\Omega(x) \tag{2.1.5}$$

that satisfies a refinement equation of the type (2.1.3). Note that this is a direct analogy to the separable HAAR scaling function. There one looks for the characteristic function of the unit cube $[0,1]^d$. For the non-separable function (2.1.5) the set Ω is covered by dilated and translated versions of itself. This opens a connection to the theory of self-affine tilings which is carried out in Gröchenig and Madych (1992). While for the scaling matrix $2\mathcal{I}$ the set Q can be shown immediately it is not obvious how this set looks like for the QUINCUNX matrix. In this case the generalization of the HAAR

[1]Speak [kwink$\Lambda\eta$ks]! *Five objects arranged so that four are at the corners of a square or rectangle and the fifth is at its centre.* Take a look at a dice, or the five of a suit of cards. In each case the dots or pips are arranged in this distinctive shape. The word comes to us from Latin, in which it literally means „five twelfths“, from *quinque*, five, plus *uncia*, a twelfth. The latter word, by the way, is also the source of our inch and of ounce (there are sixteen ounces to the pound that is used in some countries today, but that is a medieval innovation - the troy pound employed for precious metals and gems keeps the older twelve ounces). The Romans used quincunx as a symbol or marker for five-twelfths of an as, the latter being a Roman copper coin which at one time weighed twelve ounces (which could be classed as an item of small change only if you are halfway to being a giant). Learned Englishmen brought it into the language in the seventeenth century to refer to things arranged in this characteristic way. An early user was Sir Thomas Browne, in his Garden of Cyrus of 1658; this is a work of fantasy in which he traces the history of horticulture down to the time of the Persian King Cyrus. The king is credited with having been the first to plant trees in a quincunx, though Browne claimed to have discovered that it also appeared in the hanging gardens of Babylon. The diarist John Evelyn soon followed Sir Thomas's lead-in his book on orcharding, Pomona, he suggested it was a convenient way to lay out apple or pear trees. At about the same period, quincunx began to be used in astrology to refer to an aspect of planets that are five signs of the zodiac apart (out of the twelve).
The Galton board is sometimes also known as the quincunx.
If you need the adjective (although hardly anyone ever does), it is quincuncial.
Taken from `http://www.worldwidewords.org`.

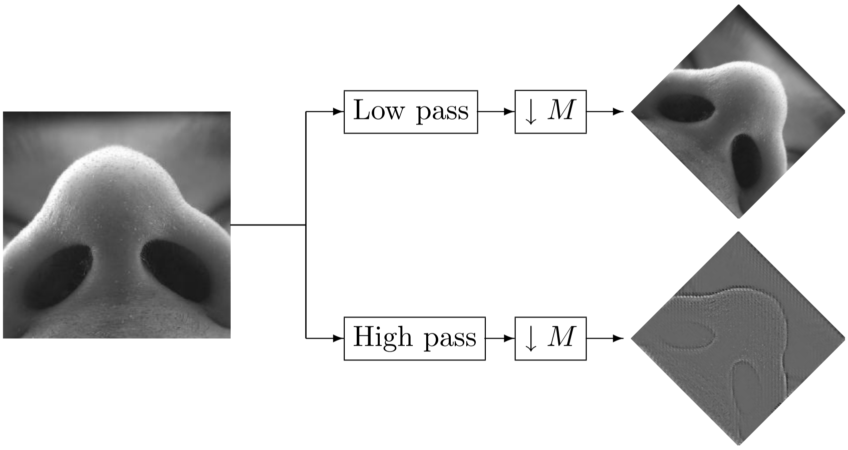


Figure 2.1: Demonstration of QUINCUNX downsampling. Images are transformed by a rotation of 45° and a dilation by factor $\sqrt{2}$.

function leads to the indicator function of a fractal set which is called *twin dragon*, see Figure 2.2.

A nice feature of the QUINCUNX matrix is that the subsampling factor is twice as fine as for the dyadic separable decomposition. Namely, its determinant is 2 and its eigenvalues have modulus $|\lambda_1| = |\lambda_2| = \sqrt{2}$ in comparison to subsampling factor 2 for the dyadic decomposition (cf. Figure 2.3). This allows a finer analysis of small textured images (cf. Figure 2.5). For a more detailed discussion the reader is referred to Mojsilovic and Popovic (1998).

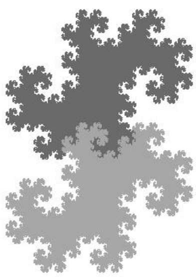


Figure 2.2: The fractal twin dragon set as a generalization of the HAAR function. Two small twin dragons are rotated by 45° and dilated by factor $\sqrt{2}$. Then they are arranged to build a self-similar larger version of the small ones.

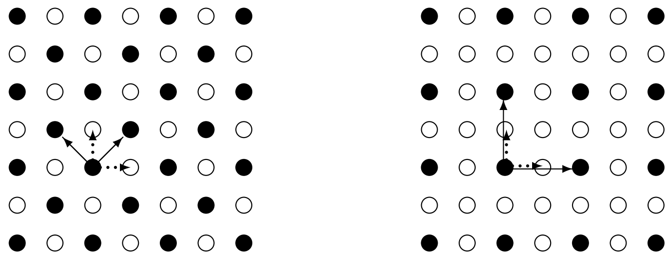


Figure 2.3: The subsampling grids: QUINCUNX downsampling (left) and separable dyadic subsampling by factor two (right).

Now we present a general definition of multivariate multiresolution decompositions which also includes separable and nonseparable settings.

Definition 2.1.2 (Multiresolution Analysis). *A multiresolution analysis of $L_2(\mathbb{R}^d)$ is a nested sequence $\{V_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L_2(\mathbb{R}^d)$ with*

$$\cdots \subset V_j \subset V_{j+1} \subset V_{j+2} \subset \cdots , \qquad j \in \mathbb{Z},$$

whose union is dense in $L_2(\mathbb{R}^d)$

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L_2(\mathbb{R}^d), \tag{2.1.6}$$

and intersection is zero

$$\bigcap_{j=-\infty}^{\infty} V_j = \{0\}. \tag{2.1.7}$$

Furthermore, each space V_j is spanned by a set of functions

$$\Phi_j := \{\phi_{j,k}, k \in \mathbb{Z}^d\} \tag{2.1.8}$$

with countable index set ∇_j. That is

$$V_j = \overline{\operatorname{span}\{\phi_{j,k}, k \in \mathbb{Z}^d\}}, \quad \phi_{j,k} \in L_2(\mathbb{R}^d)\,, \tag{2.1.9}$$

where we comprehend the closure of all finite linear combinations of elements of Φ_j in L_2 sense.
Finally, the family $\phi(\cdot - k), k \in \mathbb{Z}^d$ forms a RIESZ *basis of V_0.*

Remark 2.1.1. *We recall that a family $\{\phi_k\}_{k\in\mathbb{Z}^d}$ is called a* Riesz *basis of a Hilbert space H if and only if it spans H and the set of all finite linear combinations of ϕ_k is dense in H. Also, one has the following inequality*

$$A\sum_k |c_k|^2 \le \|\sum_k c_k\phi_k\|_H^2 \le B\sum_k |c_k|^2 \tag{2.1.10}$$

for positive constants A and B which describes a stability condition which is accomplished in more detail in Section 2.3.

As the definition of a multiresolution analysis motivates we call the function $\phi \in V_0$ the *generator* (or *scaling function*) of the multiresolution analysis. A collection of properties of a multiresolution analysis is worked out in de Boor, DeVore, and Ron (1994) and de Boor, DeVore, and Ron (1993). A general overview with many examples can be found in Wojtaszczyk (1997, Chap.5).

A multiresolution is always associated with a scaling matrix M. We call the multiresolution analysis *shift-invariant* if

$$f(\cdot) \in V_j \text{ if and only if } f(\cdot - M^{-j}k) \in V_j \text{ for all } k \in \mathbb{Z}^d. \tag{2.1.11}$$

A shift-invariant multiresolution analysis is called *stationary* if

$$f(\cdot) \in V_j \text{ if and only if } f(M\cdot) \in V_{j+1}. \tag{2.1.12}$$

A multiresolution analysis is *orthonormal* if the functions from Φ_j satisfy the condition

$$\langle \phi_{j,k}, \phi_{j,k'}\rangle = \delta_{k,k'}, \qquad k, k' \in \mathbb{Z}^d. \tag{2.1.13}$$

A stationary multiresolution analysis is called *classical* if

$$\phi_{j,k}(\cdot) := |\det M|^{j/2}\phi(M^j\cdot - k), \quad k \in \mathbb{Z}^d, j \in \mathbb{Z}, \quad \phi \in L_2(\mathbb{R}^d). \tag{2.1.14}$$

In what follows we will always assume that ϕ is normalized in $L_2(\mathbb{R}^d)$, i.e.

$$\|\phi\|_{L_2(\mathbb{R}^d)} = 1. \tag{2.1.15}$$

Thus for general L_p norms we get for all $k \in \mathbb{Z}^d$

$$\|\phi_{j,k}\|_{L_p(\mathbb{R}^d)} = |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\|\phi\|_{L_p(\mathbb{R}^d)} \qquad \text{for } 1 \le p < \infty \tag{2.1.16}$$

and

$$\|\phi_{j,k}\|_{L_\infty(\mathbb{R}^d)} = |\det M|^{j/2}\|\phi\|_{L_\infty(\mathbb{R}^d)}. \tag{2.1.17}$$

Remark 2.1.2. *Since the spaces V_j are nested it follows by condition (2.1.9) and (2.1.12) that the function ϕ satisfies the refinement equation (2.1.3) if the function ϕ is ℓ_2-stable. On the other hand refinable functions lead to a multiresolution analysis under very mild conditions (see e.g., de Boor, DeVore, and Ron (1993), or Jia and Micchelli (1991)).*

An important benefit of a multiresolution analysis is the facility to construct wavelets. Generally a wavelet is a function that is considered to be associated with a basis for L_p-spaces. A wavelet basis is generated by a finite system $\{\psi_e\}_{e\in E}$ of so called *mother wavelets.* By dilation and translation they produce the elements of the basis. If the system

$$\psi_{e,j,k}(\cdot) := |\det M|^{j/2}\psi_e(M^j \cdot -k), \qquad e \in E, j \in \mathbb{Z}, k \in \mathbb{Z}^d \tag{2.1.18}$$

generates the span of $L_2(\mathbb{R}^d)$, then $\{\psi_{e,j,k}\}_{e\in E,j\in\mathbb{Z},k\in\mathbb{Z}^d}$ is called a *wavelet basis.*

Classically, this basis was introduced as an orthonormal basis which turned out to be too restrictive for many applications. A suitable generalization with more flexibility is the concept of a *biorthogonal* wavelet basis. There one has two families $\{\psi_{e,j,k}\}$ and $\{\tilde{\psi}_{e',j',k'}\}$ which satisfy the biorthogonality condition

$$\langle \psi_{e,j,k}, \tilde{\psi}_{e',j',k'} \rangle = \delta_{e,e'}\delta_{j,j'}\delta_{k,k'}. \tag{2.1.19}$$

The main advantage for this setting is the much higher flexibility for the construction of wavelets with unchanging computational aspects. More detailed descriptions can be found in Cohen, Daubechies, and Feauveau (1992) or Cohen and Daubechies (1993).

The construction of a wavelet basis consists of finding a system $\{\psi_e\}_{e\in E}$ of functions that generates the orthogonal complement spaces W_1^e of V_0 in V_1 such that

$$V_1 = V_0 \oplus \bigoplus_{e\in E} W_1^e, \text{ with } W_1^e = \overline{\operatorname{span}\{\psi_e(\cdot - k),\ e \in E, k \in \mathbb{Z}^d\}}. \tag{2.1.20}$$

By a rescaling argument we define

$$W_{j+1}^e := \{f \in L_2(\mathbb{R}^d) |\ f(M^{-j}\cdot) \in W_1^e\} \tag{2.1.21}$$

and obtain that

$$V_{j+1} = V_j \oplus W_{j+1}^e, \qquad e \in E. \tag{2.1.22}$$

By the properties (2.1.6) and (2.1.7) of a multiresolution analysis we get a full decomposition of the space $L_2(\mathbb{R}^d)$ in terms of a wavelet basis

$$L_2(\mathbb{R}^d) = \bigoplus_{j=-\infty}^{\infty} \bigoplus_{e \in E} W_j^e. \tag{2.1.23}$$

Again, using the concept of biorthogonal systems this can be adopted to two sequences of nested subspaces $\{V_j\}_{j\in\mathbb{Z}}$, $\{\tilde{V}_j\}_{j\in\mathbb{Z}}$ of $L_2(\mathbb{R}^d)$ which both have to satisfy the conditions (2.1.6), (2.1.7), (2.1.9) and (2.1.11). The orthogonality relation (2.1.20) is replaced by

$$V_0 \perp \bigoplus_{e\in E} \tilde{W}_1^e, \quad \tilde{V}_0 \perp \bigoplus_{e\in E} W_1^e \tag{2.1.24}$$

such that

$$V_1 = V_0 \oplus \bigoplus_{e\in E} W_1^e, \quad \tilde{V}_1 = \tilde{V}_0 \oplus \bigoplus_{e\in E} \tilde{W}_1^e. \tag{2.1.25}$$

For more details the reader is referred to Cohen, Daubechies, and Feauveau (1992).

Remark 2.1.3. *A surprising fact is that the number of different wavelets from the set E for each level depends on the scaling matrix. Since the scaling matrix M is an integer matrix it follows $M(\mathbb{Z}^d) \subset \mathbb{Z}^d$. Also, since we claim that M is expanding one has $|\det M| > 1$. If we consider $\mathbb{Z}^d$ as an additive group then $M(\mathbb{Z}^d)$ is a normal subgroup. The cosets of $M(\mathbb{Z}^d)$ in $\mathbb{Z}^d$ also form a group. It can be shown that the number of different cosets equals $|\det M|$ which determines the number of wavelets to span the complement spaces for each level as $|E| = |\det M| - 1$ since we have to subtract one for the scaling function. A nice proof of this fact can be found e.g. in Wojtaszczyk (1997, Chap.5).*

For a 1-D dyadic multiresolution analysis with scaling factor two one always has one scaling function and one wavelet. Similarly for a 1-D dyadic multiresolution analysis with a different scaling factor s than two, one needs $s-1$ wavelets to build W_0.

For the tensor product case in two dimensions $d = 2$ the scaling matrix $M = 2\mathcal{I}$ has determinant $|\det M| = 4$. This implies that the scaling function ϕ is associated with $|\det M| - 1 = 3$ wavelets. These are

$$\psi(x_1)\psi(x_2), \qquad \psi(x_1)\phi(x_2), \qquad \phi(x_1)\psi(x_2).$$

As it is known that multivariate scaling functions can be obtained easily via tensor products we have also seen that already in two dimensions we need three wavelets.

From the viewpoint of application it is important to keep the amount of computations small and therefore a good choice is a scaling matrix M with $|\det M| = 2$. This has also another consequence: The choice to keep $|\det M|$ as small as possible also keeps the eigenvalues of M small. This results as we noted before in a finer resolution between levels. In Figure 2.4

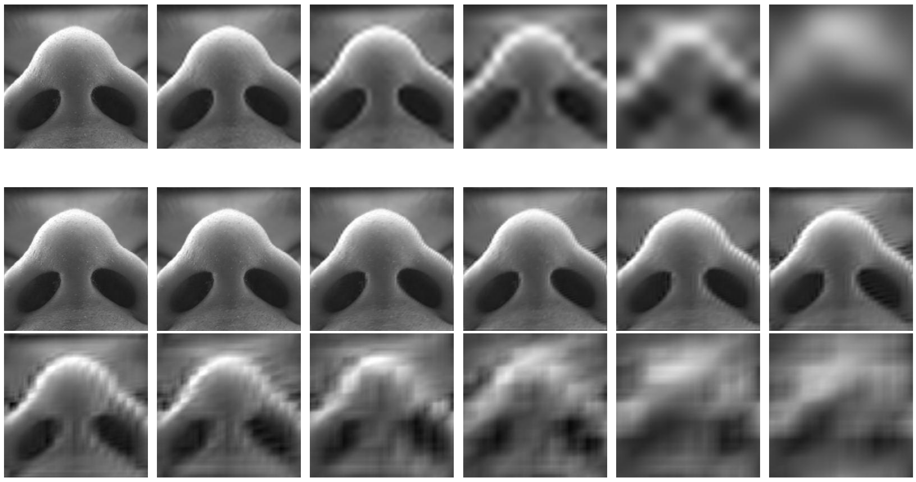


Figure 2.4: Top row: Reconstruction from truncated six-level decomposition with separable bior3.5 wavelet. Truncations of wavelet coefficients from top tree level 1 to 6, i.e. these levels are set to zero. Bottom rows: Reconstruction from truncated sixteen-level decomposition with non-separable „Resting Dog" wavelet (cf. Belogay and Wang 1999). Truncations of wavelet coefficients from top tree level 1 to 12. (Software for non-separable transform written by F. Mendivil (Acadia University, Canada))

we illustrate that the first and last reconstructions from truncated wavelet expansions appear to be similar. At the same time we get visually finer distinctions in quality for a non-separable wavelet decomposition with the continuous (not differentiable) „Resting Dog" wavelet. This is associated with the scaling matrix $\begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$ with the coefficient mask given in Table 2.1 (cf. Belogay and Wang (1999)). In comparison to that we depict in Figure 2.5 the corresponding reconstructions for truncated wavelet decompositions from the bottom keeping the leading approximation coefficients. Again, we see that the fine structures are better resolved with the non-separable decomposition due to the finer dilation factor.

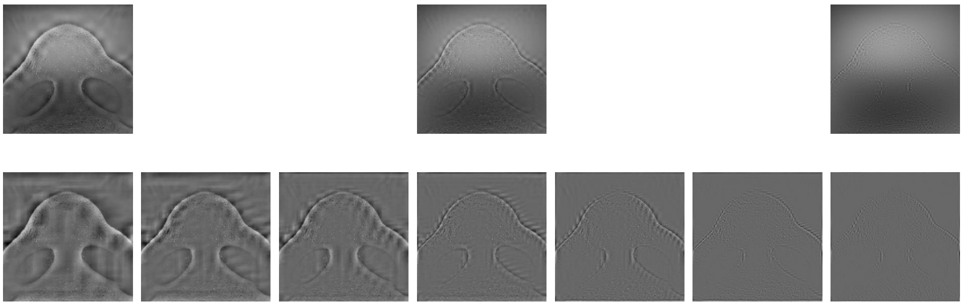


Figure 2.5: Top row: Reconstruction from truncated six-level decomposition with separable bior3.5 wavelet. Truncations of wavelet coefficients from bottom tree level 3 to 5, i.e. these levels are set to zero. Bottom row: Reconstruction from truncated sixteen-level decomposition with non-separable „Resting Dog“ wavelet . Truncations of wavelet coefficients from bottom tree level 7 to 13. (Software for non-separable transform written by F. Mendivil (Acadia University, Canada))

To avoid the preference of specific directions the use of isotropic wavelets is desirable. We have already mentioned that separable wavelet decompositions exhibit a strong anisotropy for horizontal, vertical and diagonal directions. Non-separable decompositions feature a less directional resolution.

Another directional feature that we introduce now is the isotropy for the scaling matrix. This describes that the local refinement process is isotropic, that is the scaling matrix dilates equally in all directions.

Definition 2.1.3 (Isotropic). *We say an $d \times d$ scaling matrix M with entries in $\mathbb{Z}$ is* isotropic *if M is similar to a diagonal matrix* $\operatorname{diag}\{\lambda_1, \ldots, \lambda_d\}$ *with* $|\lambda_1| = \ldots = |\lambda_d|$.

The property of isotropic dilation should be kept also for non-separable wavelet decompositions. The basic requisites work for isotropic scaling matrices. Therefore we mainly consider this setting in the sequel. It is useful to introduce a short notation for the largest dilation of M. We mean by

$$\rho = \rho(M) = \max_{i=1,\ldots,d} |\lambda_i| \tag{2.1.26}$$

with λ_i denoting the eigenvalues of the matrix M.

Note that we distinct here isotropy for the scaling and isotropy for the analysis direction of the wavelets. There exist some works about wavelet characterizations for anisotropic Besov spaces, e.g. Garrigós and Tabacco (2002), Garrigós, Hochmuth, and Tabacco (2004) and Hochmuth (2002b) or

locations first half	first half	second half
0 0	.01981170613	.01580714174
1 0	-.03431488160	-.02737877262
2 0	-.1676882939	-.2761680314
3 0	-.005308530658	.2423648376
4 0	.2759414694	1.099852188
5 0	.1135617061	.6923441928
6 0	-.1280648816	.1605087019
7 0	-.07393829386	.09266974222
0 1	.09266974222	.07393829386
1 1	-.1605087019	-.1280648816
2 1	.6923441928	-.1135617061
3 1	-1.099852188	.2759414694
4 1	.2423648376	.005308530658
5 1	.2761680314	-.1676882939
6 1	-.02737877262	.03431488160
7 1	-.01580714174	.01981170613

Table 2.1: Coefficient mask for the „Resting dog“ wavelet. First column shows the locations of the first half of the filter coefficients. The second half of the locations is obtained by symmetry.

Hochmuth (2002a). In this thesis we emphasize the isotropy of the scaling matrix. Indeed, it is also possible to have anisotropic wavelet bases with isotropic dilation matrices as it is the case for the dyadic multiresolution analysis.

Example 2.1.1. *The* Quincunx *matrix* $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ *has determinant 2 and modulus of eigenvalues* $|\lambda_1| = |\lambda_2| = \sqrt{2}$ *and is thus isotropic.*

In general, all matrices of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ *with* $a, b \in \mathbb{R}$ *are isotropic.*

Example 2.1.2. *The matrices* $\begin{pmatrix} 0 & -2 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$ *and* $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ *have determinant* $2, -2, -2$ *and modulus of eigenvalues* $|\lambda_1| = |\lambda_2| = \sqrt{2}$ *and are also isotropic. Moreover, for these matrices one has* $M^2 = \pm 2\mathcal{I}$.

Example 2.1.3. *The matrix* $\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$ *has determinant* 3 *and modulus of eigenvalues* $|\lambda_1| = |\lambda_2| = \sqrt{3}$ *and is isotropic. In Figure 2.6 we depict a* *decomposition with reconstruction from the truncated wavelet decomposition. The associated coefficient mask for the scaling functions is*

$$(1.41421356, -0.70710678, -0.70710678, 0.0000000, 1.224744871, -1.224744871)$$

and the locations are $(0,0)$*,* $(0,1)$ *and* $(-1,0)$ *where we only put one half of the location since the mask is symmetric.*

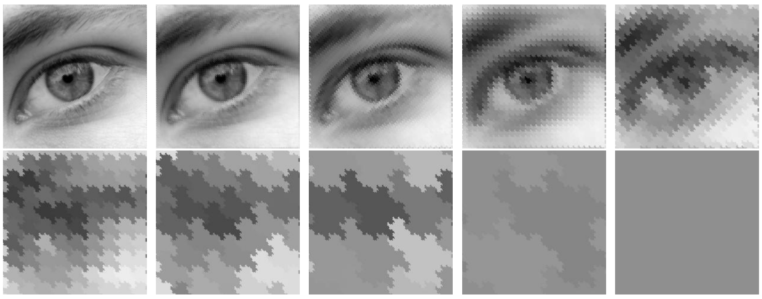


Figure 2.6: Reconstruction from truncated ten-level decomposition with non-separable wavelet with scaling matrix $M = [1-2; 2-1]$. Truncations of wavelet coefficients from top tree level 1 to 10, i.e. these levels are set to zero. (Software for non-separable transform written by F. Mendivil (Acadia University, Canada))

Remark 2.1.4. *We recall that by the inequality of* HADAMARD

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & & \vdots \\ a_{d1} & \cdots & a_{dd} \end{pmatrix}^2 \leq \sum_{k=1}^{d} |a_{k1}|^2 \cdots \sum_{k=1}^{d} |a_{kd}|^2 \tag{2.1.27}$$

we obtain in particular that $|\det M| \leq \|M\|_2^d$ *which implies that* $|\det M^{-j}| \leq \|M^{-j}\|_2^d$*. By* $\|M\|_2$ *we denote the spectral norm of the matrix* M *which coincides for isotropic scaling matrices with* ρ *defined in (2.1.26).*

As an introduction into the concept of a non-separable multiresolution analysis we have seen the joint properties as well as significant differences to separable decompositions. In the next section we introduce some useful notations that we need later on.

2.2 Some useful Notations

In Section 2.3 we turn back to the scaling function basis as defined in (2.1.8) to have a deeper insight in its stability properties. First we introduce some useful notations. These conventions are frequently used in the wavelet community.

For the projections on the spaces V_j and $\tilde{V}_j$ we write

$$P_j f = \sum_{k\in\mathbb{Z}^d} \langle f, \tilde{\phi}_{j,k}\rangle \phi_{j,k} \quad \text{and} \quad \tilde{P}_j f = \sum_{k\in\mathbb{Z}^d} \langle f, \phi_{j,k}\rangle \tilde{\phi}_{j,k}, \tag{2.2.1}$$

where $\phi_{j,k}$ is the scaling function for V_j and $\tilde{\phi}_{j,k}$ is the corresponding dual function. These projections are adjoint due to the duality principle and the biorthogonality relation. One easily checks that

$$P_j P_n = P_j, \qquad \text{for } j \le n. \tag{2.2.2}$$

Similarly, we introduce the projections on the complement spaces W_j^e and $\tilde{W}_j^e$ as

$$Q_j f = \sum_{e\in E}\sum_{k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k} \quad \text{and} \quad \tilde{Q}_j f = \sum_{e\in E}\sum_{k\in\mathbb{Z}^d} \tilde{d}_{e,j,k}\tilde{\psi}_{e,j,k}, \tag{2.2.3}$$

where $d_{e,j,k} = \langle f, \tilde{\psi}_{e,j,k}\rangle$ and $\tilde{d}_{e,j,k} = \langle f, \psi_{e,j,k}\rangle$.

The representation of functions f in V_{j+1} as in (2.1.22) implies a total decomposition of V_m as follows

$$V_m = \bigoplus_{j=j_0}^{m} \bigoplus_{e\in E} W_j^e \quad \text{with } W_{j_0}^e := V_{j_0}, \tag{2.2.4}$$

assuming that $j = j_0$ is the coarsest level such that $f \in V_j$. In what follows we will always suppose that $j_0 = 0$. All other cases follow immediately by rescaling.

Each function $g \in V_m$ then has a *multilevel representation*

$$g = P_0 g + \sum_{j=1}^{m} (P_j - P_{j-1})g. \tag{2.2.5}$$

Defining $\Psi_0 := \Phi_0$, $W_0^e := V_0$ and $\Psi_j := \{\psi_{e,j,k}, e \in E, k \in \nabla_j\}$ we obtain an alternative multilevel representation for a function $g \in V_m$ by

$$g = \sum_{e\in E}\sum_{j=0}^{m}\sum_{k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k}. \tag{2.2.6}$$

The same properties are valid for the dual spaces $\{\tilde{V}_j\}_{j\in\mathbb{Z}}$.

Further on, we will write $W_j := \bigoplus_{e\in E} W_j^e$.

It is useful to introduce the abbreviation $\lambda = (e, j, k)$ for the indices of wavelet coefficients $d_{e,j,k}$ associated with a wavelet decomposition of a function $f = \sum_{e\in E, j\in\mathbb{Z}, k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k}$. We define for these indices

$$|\lambda| = |(e, j, k)| := j. \tag{2.2.7}$$

When we consider wavelet coefficients we denote the set of all indices of the wavelet coefficients by

$$\nabla := \nabla(f) = \{\lambda = (e, j, k) : \ e \in E, j \in \mathbb{Z}, k \in \mathbb{Z}^d\}. \tag{2.2.8}$$

For the set of all indices of the wavelet coefficients on level j we write

$$\nabla_j := \{\lambda = (e, j, k) \in \nabla : \ |\lambda| = j, e \in E, k \in \mathbb{Z}^d\}. \tag{2.2.9}$$

The set of all wavelet coefficients on level j is denoted by

$$d^{(j)} := \{d_\lambda : \ \lambda \in \nabla_j\}. \tag{2.2.10}$$

We denote by $\mathbf{T}_n$ the transformation which takes the coefficients $d^{(j)}$ of the multilevel representation into the coefficients of the single scale representation, the so-called *nodal basis* with coefficients c_k. For numerical computations the transformation $\mathbf{T}_n$ plays an important role. This transformation should be efficiently executable which can be obtained by sufficiently localized basis functions, see Dahmen (1996) for more details. For accurate computations it is necessary that $\mathbf{T}_n$ is well conditioned, that is

$$\|\mathbf{T}_n\|_2 \cdot \|\mathbf{T}_n^{-1}\|_2 = \mathcal{O}(1), \qquad n \to \infty, \tag{2.2.11}$$

where $\|\cdot\|_2$ denotes the spectral norm. Also in Dahmen (1996) it is shown that the operators $\mathbf{T}_n$ are well conditioned in the sense of (2.2.11) if and only if $\Psi := \bigcup_{j=0}^{\infty} \Psi_j$ is a RIESZ basis of $L_2(\mathbb{R}^d)$. Equivalent for $\mathbf{T}_n$ being well conditioned is that all functions $f \in L_2(\mathbb{R}^d)$ have a unique representation

$$f = \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k} \tag{2.2.12}$$

and the basis Ψ is stable in the sense of (2.3.1).

In the next section we examine stability properties of non-separable wavelet decompositions.

2.3 Orthogonality and Stability

Later on we will see that stability properties are of central meaning for the identification of a given function by its coefficient sequence. Therefore we are interested in simple criteria by which we can check the following stability condition:

Definition 2.3.1 (Stable Multiresolution Analysis). *A multiresolution analysis is said to be* stable *if for all j the basis Φ_j is uniformly ℓ_2-stable, that is,*

$$\left\| \sum_{k\in\mathbb{Z}^d} c_k \phi_{j,k} \right\|_{L_2(\mathbb{R}^d)} \sim \|(c_k)_{k\in\mathbb{Z}^d}\|_{\ell_2(\mathbb{Z}^d)}. \tag{2.3.1}$$

The notation ‚$a \sim b$‘ means that both a and b can be uniformly bounded by a constant multiple of each other, that is there exist positive constants C_1, C_2 such that

$$C_1 b \leq a \leq C_2 b. \tag{2.3.2}$$

Likewise, we write ‚$a \lesssim b$‘ (or ‚$a \gtrsim b$‘ resp.) if there exists a positive constant C such that $a \leq Cb$ (or $a \geq Cb$ resp.).

Numerical computations for functions are usually done for the coefficients of their basis expansions rather than for the functions themselves. If we have in hand a stability criterion for such a basis we make sure that small approximation errors for the coefficients lead to small changes for the resulting approximated function.

The stability property (2.3.1) also implies that the mapping

$$(c_{j,k})_{k\in\mathbb{Z}^d} \longmapsto \sum_{k\in\mathbb{Z}^d} c_{j,k}\phi_{j,k}$$

generates an isomorphism from $\ell_2(\mathbb{Z}^d)$ to $L_2(\mathbb{R}^d)$. Furthermore, the series $\sum_{k\in\mathbb{Z}^d} c_{j,k}\phi_{j,k}$ converges unconditionally in $L_2(\mathbb{R}^d)$, i.e. the terms of the sum can be arranged in any order without affecting the convergence if and only if $\sum_{k\in\mathbb{Z}^d} |c_{j,k}|^2 < \infty$.

Another important fact is that any $f \in L_2(\mathbb{R}^d)$ can be uniquely decomposed in a linear combination of the $\phi_{j,k}$. Finally, there exists a corresponding unique dual basis $\{\tilde{\phi}_{j,k}\}_{j\in\mathbb{Z},k\in\mathbb{Z}^d}$ such that $\langle \phi_{j,k}, \tilde{\phi}_{j',k'}\rangle = \delta_{j,j'}\delta_{k,k'}$. Then the expansion coefficients are given by $c_{j,k} = \langle f, \tilde{\phi}_{j,k}\rangle$ such that

$$f = \sum_{j,k} \langle f, \tilde{\phi}_{j,k}\rangle \phi_{j,k} = \sum_{j,k} \langle f, \phi_{j,k}\rangle \tilde{\phi}_{j,k}. \tag{2.3.3}$$

The aim of this section is to provide criteria that relate properties of the scaling function basis with their stability. First we confine our considerations to suitable subspaces of $L_2(\mathbb{R}^d)$ (or $L_p(\mathbb{R}^d)$, respectively). In many practical applications, this restriction is insubstantial since these subspaces are in general large enough to describe a wide class of signals or images.

For any function f (compactly supported or rapidly decaying) we define the periodization operator $[f]$ by

$$[f] := \sum_{k \in \mathbb{Z}^d} |f(\cdot - k)|. \tag{2.3.4}$$

A suitable norm for these functions is

$$|f|_p := \|[f]\|_{L_p([0,1)^d)}, \quad 1 \leq p \leq \infty. \tag{2.3.5}$$

We proceed by introducing the subspace $\mathcal{L}_p(\mathbb{R}^d)$ of $L_1(\mathbb{R}^d)$.

Definition 2.3.2 ($\mathcal{L}_p(\mathbb{R}^d)$ Spaces). *The space $\mathcal{L}_p(\mathbb{R}^d)$ is the* Banach *space*

$$\mathcal{L}_p(\mathbb{R}^d) := \{f : \ |f|_p < \infty\}, \quad 1 \leq p \leq \infty \tag{2.3.6}$$

To get a bit more familiar with the spaces $\mathcal{L}_p(\mathbb{R}^d)$ we record some facts about them.

Theorem 2.3.1. *(i) $\mathcal{L}_p(\mathbb{R}^d) \subset L_p(\mathbb{R}^d)$ for $1 \leq p \leq \infty$.*

(ii) $\mathcal{L}_p(\mathbb{R}^d) \subset \mathcal{L}_q(\mathbb{R}^d)$ for $1 \leq q < p \leq \infty$.

(iii) If $|f(x)| \lesssim (1 + |x|)^{-d-\delta}$ for all $x \in \mathbb{R}^d$ and positive $\delta \in \mathbb{R}$, then $f \in \mathcal{L}_\infty(\mathbb{R}^d)$.

Proof. To show (i) we derive

$$\begin{aligned}
\|f\|_p &= \left(\int_{\mathbb{R}^d} |f(x)|^p \mathrm{d}x\right)^{\frac{1}{p}} \\
&= \left(\sum_{\alpha\in\mathbb{Z}^d} \int_{[0,1]^d+\alpha} |f(x)|^p \mathrm{d}x\right)^{\frac{1}{p}} \\
&= \left(\sum_{\alpha\in\mathbb{Z}^d} \int_{[0,1]^d} |f(x-\alpha)|^p \mathrm{d}x\right)^{\frac{1}{p}} \\
&= \left(\int_{[0,1]^d} \sum_{\alpha\in\mathbb{Z}^d} |f(x-\alpha)|^p \mathrm{d}x\right)^{\frac{1}{p}} \\
&\geq \left(\int_{[0,1]^d} \left(\sum_{\alpha\in\mathbb{Z}^d} |f(x-\alpha)|\right)^p \mathrm{d}x\right)^{\frac{1}{p}} = |f|_p.
\end{aligned}$$

Assume that $1 \leq p \leq \infty$ and $f \in \mathcal{L}_p(\mathbb{R}^d)$. Then assertion (ii) can be obtained by HÖLDER's inequality since $\frac{1}{p/q} + \frac{1}{p/(p-q)} = 1$.

$$\begin{aligned}
|f|_q &= \left(\int_{[0,1]^d} (\sum_{\alpha\in\mathbb{Z}^d} |f(x-\alpha)|)^q \mathrm{d}x\right)^{\frac{1}{q}} \\
&= \left\langle (\sum_{\alpha\in\mathbb{Z}^d} |f(x-\alpha)|)^q, 1 \right\rangle_{L_2([0,1]^d)}^{\frac{1}{q}} \\
&\leq \left\| (\sum_{\alpha\in\mathbb{Z}^d} |f(x-\alpha)|)^q \right\|_{L_{\frac{p}{q}}([0,1]^d)}^{\frac{1}{q}} \|1\|_{L_{p/(p-q)}([0,1]^d)}^{\frac{1}{q}} \\
&= \left(\int_{[0,1]^d} (\sum_{\alpha\in\mathbb{Z}^d} |f(x-\alpha)|)^{p\frac{p}{q}} \mathrm{d}x\right)^{\frac{q}{p}\frac{1}{q}} = |f|_p.
\end{aligned}$$

For the third proposition we see that it is sufficient to show

$$|f|_\infty \leq \operatorname{ess}\sup_{x\in[0,1]^d} \sum_{\alpha\in\mathbb{Z}^d} (1+|x-\alpha|)^{-d-\delta} < \infty.$$

To prove this inequality we introduce polar coordinates

$$\begin{aligned}\int_{\mathbb{R}^d}(1+|x-y|)^{-d-\delta}\mathrm{d}y &= \int_{\mathbb{R}^d}(1+|y|)^{-d-\delta}\mathrm{d}y\\ &\sim \int_{\mathbb{R}}(1+r)^{-d-\delta}r^{d-1}\mathrm{d}r\\ &= \int_{\mathbb{R}}\frac{r^{d-1}}{(1+r)^{d-1}}\frac{1}{(1+r)^{1+\delta}}\mathrm{d}r\\ &\lesssim \int_{\mathbb{R}}\frac{1}{(1+r)^{1+\delta}}\mathrm{d}r.\end{aligned}$$

An antiderivative of the integrand is $\frac{1}{\delta}\frac{1}{(1+r)^{1+\delta}}$. Since $\delta > 0$ we conclude that the integral exists and $f \in \mathcal{L}_\infty(\mathbb{R}^d)$. □

In the sequel we make use of a short notation which is expressed as a convolution type operation. We introduce the so called *semi-discrete convolution.*

Definition 2.3.3 (semi-discrete convolution). *For a function $\phi \in \mathcal{L}_p(\mathbb{R}^d)$ and a sequence $\mathbf{a} \in \ell_\infty(\mathbb{Z}^d)$ we set*

$$(\phi *' \mathbf{a}) := \sum_{k\in\mathbb{Z}^d} a_k \phi(\cdot - k). \tag{2.3.7}$$

An amazing fact for the above defined $\mathcal{L}_p$-spaces is that for a stability condition it suffices to show the lower bound of (2.3.1). The upper bound always exists. This connection was carried out e.g. in Jia and Micchelli (1991, Th.2.1).

Definition 2.3.4 (ℓ_p-stable). *A function $\phi \in \mathcal{L}_p(\mathbb{R}^d)$ is called* ℓ_p-stable *if for all $\mathbf{a} \in \ell_p(\mathbb{Z}^d)$*

$$\|\phi *' \mathbf{a}\|_p \gtrsim \|\mathbf{a}\|_{\ell_p}. \tag{2.3.8}$$

Remark 2.3.1. *In Definition 2.3.4 we require in contrary to (2.3.1) just a bound from below. This makes sense since in the setting of $\mathcal{L}_p$-spaces the boundedness from above is always valid as the following theorem shows.*

Theorem 2.3.2. *For $\phi \in \mathcal{L}_p(\mathbb{R}^d)$ and $\mathbf{a} \in \ell_p(\mathbb{Z}^d)$ we have the* Young *type inequalities*

$$|\phi *' \mathbf{a}|_p \leq |\phi|_p \|\mathbf{a}\|_{\ell_1}; \tag{2.3.9}$$

$$\|\phi *' \mathbf{a}\|_{L_p} \leq |\phi|_p \|\mathbf{a}\|_{\ell_p}. \tag{2.3.10}$$

Proof. First of all we show (2.3.9). Inserting (2.3.7) in (2.3.4) yields to

$$
\begin{aligned}
[(\phi *' \mathbf{a})] &= \sum_{k\in\mathbb{Z}^d} \left| \sum_{l\in\mathbb{Z}^d} a_l \phi(\cdot - l - k) \right| && (2.3.11)\\
&\leq \sum_{k\in\mathbb{Z}^d} \sum_{l\in\mathbb{Z}^d} |a_l|\, |\phi(\cdot - l - k)| && (2.3.12)\\
&\leq \|\mathbf{a}\|_{\ell_1} [\phi]. && (2.3.13)
\end{aligned}
$$

Application of the $|\cdot|_p$-norm proves the first assertion (2.3.9). It remains to show (2.3.10). We split the range of integration for the L_p-norm into integer intervals

$$
\|\phi *' \mathbf{a}\|_{L_p}^p = \sum_{l\in\mathbb{Z}^d} \int_{[0,1)^d+l} |(\phi *' \mathbf{a})(x)|^p \,\mathrm{d}x = \int_{[0,1)^d} \sum_{l\in\mathbb{Z}^d} |(\phi *' \mathbf{a})(x+l)|^p \,\mathrm{d}x.
$$

For a fixed $x \in \mathbb{R}^d$ we define the sequence $\mathbf{c} := (\phi(x+n))_{n\in\mathbb{Z}^d}$. Then we rewrite the semidiscrete convolution by means of a common convolution as

$$
(\phi *' \mathbf{a})(x+l) = (\mathbf{a} * \mathbf{c})(l)
$$

and the inequality of Young implies

$$
\|\mathbf{a} * \mathbf{c}\|_{\ell_p} \leq \|\mathbf{a}\|_{\ell_p} \|\mathbf{c}\|_{\ell_1}.
$$

Hence,

$$
\begin{aligned}
\sum_{l\in\mathbb{Z}^d} |(\phi *' \mathbf{a})(x+l)|^p &= \|\mathbf{a} * \mathbf{c}\|_{\ell_p}^p && (2.3.14)\\
&\leq \|\mathbf{a}\|_{\ell_p}^p \|\mathbf{c}\|_{\ell_1}^p && (2.3.15)\\
&= \|\mathbf{a}\|_{\ell_p}^p ([\phi](x))^p. && (2.3.16)
\end{aligned}
$$

Altogether, it follows

$$
\|\phi *' \mathbf{a}\|_{L_p}^p \leq \|\mathbf{a}\|_{\ell_p}^p \int_{[0,1)^d} ([\phi](x))^p \,\mathrm{d}x = \|\mathbf{a}\|_{\ell_p}^p |\phi|_p^p.
$$

□

In practice, the most important space among the $\mathcal{L}_p$-spaces is $\mathcal{L}_2(\mathbb{R}^d)$. But even for arbitrary p there exists a nice characterization of ℓ_p-stable scaling functions by means of their Fourier transform. In the sequel we denote by $\hat{\phi}$ the Fourier transform of a function $\phi \in L_1(\mathbb{R}^d)$, which is defined by

$$
\hat{\phi}(\xi) := \int_{\mathbb{R}^d} \phi(x) e^{-ix\xi} \mathrm{d}x. \tag{2.3.17}
$$

Theorem 2.3.3. *Let $\phi \in \mathcal{L}_p(\mathbb{R}^d)$, $1 \le p \le \infty$. ϕ has ℓ_p-stable translates if and only if for all $\xi \in \mathbb{R}^d$*

$$\sup_{k \in \mathbb{Z}^d} |\hat{\phi}(\xi + 2\pi k)| > 0. \tag{2.3.18}$$

A proof of this result can be found in Jia and Micchelli (1991, p. 224).

2.4 Stability of Non-Separable Functions

This section is devoted to the investigation of stability properties of non-separable refinable functions. In the next theorems we show that under very mild conditions the basis Φ_j and Ψ_j constitute a stable basis.

Theorem 2.4.1. *Let $1 \le p \le \infty$. Assume that $\phi \in L_p(\mathbb{R}^d)$ and $\tilde{\phi} \in L_{p'}(\mathbb{R}^d)$ are compactly supported $(\mathbf{h}, M)$-refinable $((\tilde{\mathbf{h}}, M)$-refinable resp.) scaling functions with $\frac{1}{p} + \frac{1}{p'} = 1$. Then the projectors P_j are uniformly bounded in $L_p(\mathbb{R}^d)$ and the basis Φ_j of V_j is uniformly ℓ_p-stable. That is,*

$$\|P_j f\|_{L_p(\mathbb{R}^d)} \lesssim \|f\|_{L_p(\mathbb{R}^d)} \tag{2.4.1}$$

and

$$\left\| \sum_{k \in \mathbb{Z}^d} c_k \phi_{j,k} \right\|_{L_p(\mathbb{R}^d)} \sim |\det M|^{j(\frac{1}{2} - \frac{1}{p})} \|(c_k)_{k \in \mathbb{Z}^d}\|_{\ell_p}. \tag{2.4.2}$$

where the constants do not depend on j.

Proof. First we show the upper inequality. Let $I_{j,k} := \|M^{-j}\|_\infty (k + [0,1]^d)$, $k \in \mathbb{Z}^d$ denote the scaled and shifted unit cube. Since ϕ has compact support the number of indices k such that $l - k \in \operatorname{supp}\phi$ for fixed l is finite. Hence, all sequence norms on the set of these indices are equivalent. Then we have on each $I_{j,l}$

$$\begin{aligned} \left\| \sum_{k \in \mathbb{Z}^d} c_k \phi_{j,k} \right\|_{L_p(I_{j,l})} &\lesssim \sup_{k:\ l-k \,\in\, \operatorname{supp}\phi} |c_k| \ \|\phi_{j,0}\|_{L_p(\mathbb{R}^d)} \\ &\lesssim |\det M|^{j(\frac{1}{2} - \frac{1}{p})} \|(c_k)_{k \in \mathbb{Z}^d}\|_{\ell_p}. \end{aligned} \tag{2.4.3}$$

Now we show the L_p-stability of P_j. Let $f \in L_p(\mathbb{R}^d)$.

$$|\langle f, \tilde{\phi}_{j,k}\rangle| \le \|f\|_{L_p(\operatorname{supp}\tilde{\phi}_{j,k})} \|\tilde{\phi}_{j,0}\|_{L_{p'}(\operatorname{supp}\tilde{\phi}_{j,k})} \lesssim |\det M|^{j(\frac{1}{2} - \frac{1}{p'})} \|f\|_{L_p(\operatorname{supp}\tilde{\phi}_{j,k})},$$

where we used HÖLDER's inequality. Taking the p-th power and summation over $k \in \mathbb{Z}^d$ yields

$$\|(\langle f, \tilde{\phi}_{j,k}\rangle)_{k\in\mathbb{Z}^d}\|_{\ell_p(\mathbb{Z}^d)} \lesssim |\det M|^{j(\frac{1}{2}-\frac{1}{p'})}\|f\|_{L_p(\mathbb{R}^d)} \tag{2.4.4}$$

Using the upper inequality (2.4.3) we deduce

$$\|P_j f\|_{L_p(\mathbb{R}^d)} \le C|\det M|^{j(\frac{1}{2}-\frac{1}{p})}|\det M|^{j(\frac{1}{2}-\frac{1}{p'})}\|f\|_{L_p(\mathbb{R}^d)} = C\|f\|_{L_p(\mathbb{R}^d)},$$

where the constant C does not depend on j. To show the lower inequality of the equivalence (2.4.2) we employ equation (2.4.4).

$$|\det M|^{j(\frac{1}{2}-\frac{1}{p})}\|(c_k)_{k\in\mathbb{Z}^d}\|_{\ell_p(\mathbb{Z}^d)} \le C\|f\|_{L_p(\mathbb{R}^d)},$$

which is valid for functions f in V_j that have the representation

$$f = \sum_{k\in\mathbb{Z}^d} c_k\phi_{j,k}$$

with $c_k = \langle f, \tilde{\phi}_{j,k}\rangle$. □

Remark 2.4.1. *The argument for the above stability depends only on the compact support and the equivalence of all ℓ_p norms on finite index sets. Therefore the same arguments work for wavelets. For $\psi \in L_p$ and $\tilde{\psi} \in L_{p'}$ with $\frac{1}{p}+\frac{1}{p'} = 1$ and compact support we employ the projectors*

$$Q_j f = \sum_{e\in E}\sum_{k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k} \tag{2.4.5}$$

with $d_{e,j,k} = \langle f, \tilde{\psi}_{e,j,k}\rangle$ and obtain for $1 \le p \le \infty$ that the projectors Q_j are uniformly bounded in $L_p(\mathbb{R}^d)$. Moreover the basis Ψ_j is uniformly ℓ_p-stable. That is,

$$\|Q_j f\|_{L_p(\mathbb{R}^d)} \lesssim \|f\|_{L_p(\mathbb{R}^d)} \tag{2.4.6}$$

and

$$\left\|\sum_{e\in E}\sum_{k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k}\right\|_{L_p(\mathbb{R}^d)} \sim |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\|d^{(j)}\|_{\ell_p}. \tag{2.4.7}$$

for

$$d^{(j)} := \{d_\lambda :\ \lambda \in \nabla_j\}.$$

where the constants do not depend on the level $j \in \mathbb{Z}$.

We have seen that non-separable scaling functions as well as their corresponding wavelet bases provide stable bases for L_p spaces. This fundamental property ensures that analytical properties of functions can be accessed via numerical computations on the associated discrete wavelet coefficients. Furthermore, we will see in Section 5.1 that stability is a basic ingredient for the characterization of smoothness spaces via norm equivalences between discrete weighted sequences of wavelet coefficients and discrete BESOV norms. This characterization is established by the above stated stability norm equivalences.

2.5 Smoothness Spaces

Later on we relate approximation properties of shift-invariant spaces with smoothness properties of their member functions. In this section we give a basic introduction about function spaces with focus on smoothness spaces. Particular attention applies to BESOV spaces.

There are many ways to measure smoothness. A wide range of smoothness classes such as HÖLDER, LIPSCHITZ, SOBOLEV or BESOV spaces have been intensively investigated since the middle of the last century. As we will see in the next chapter BESOV spaces turn out to be the suitable spaces for our topic since they allow an exact characterization of their member functions in terms of wavelet decompositions.

Most of these spaces are linear real (or complex) normed function spaces. In the next few paragraphs we recall some definitions that are well known and can be found in standard books such as DeVore and Lorentz (1993) or Triebel (1983). Nevertheless it will now be useful to collect all notions that we need in the subsequent chapters. In particular we work with BESOV (semi-)norms which turn out to be equivalent to a wide range of other smoothness norms, cf. (2.5.2). In Chapter 7 we point out that we deal with quasi-norms. In Section 2.5.1 we study the properties of the modulus of smoothness that is a basic component of a BESOV semi-norm.

We recall that normed function spaces are equipped with a norm $\|\cdot\|$ with the following properties

$$\begin{aligned} &(i) && \|x\| = 0 \Longleftrightarrow x = 0 \\ &(ii) && \|ax\| = |a|\ \|x\|, \quad a \in \mathbb{R} \\ &(iii) && \|x+y\| \leq \|x\| + \|y\|. \end{aligned} \tag{2.5.1}$$

If the normed space is complete we call it a BANACH space. A semi-norm $|\cdot|$ fulfills (2.5.1) except condition (i). A quasi-norm $\|\cdot\|$ on a linear space fulfills (2.5.1) except that condition (iii) is replaced by

$$(iii') \quad \|x+y\| \leq C(\|x\|+\|y\|), \tag{2.5.2}$$

for a positive constant C depending on the space.

Continuous functions with compact support play an important role for numerical computations. The class of these functions is denoted by $C_0(\Omega)$ for $\Omega \subseteq \mathbb{R}^d$.

To allow averaged boundedness, there are for $0 < p \leq \infty$ the spaces $L_p(\mathbb{R}^d)$. They consist of all measurable functions f for which the following quantity is finite

$$\|f\|_p := \|f\|_{L_p(\Omega)} = \begin{cases} \left(\int_\Omega |f(x)|^p \mathrm{d}x\right)^{1/p}, & 0 < p < \infty \\ \operatorname{ess\,sup}_{x\in\Omega} |f(x)|, & p = \infty. \end{cases} \tag{2.5.3}$$

Note that (2.5.3) is only a norm for $1 \leq p \leq \infty$. The dual space of $L_p(\Omega)$ with $1 \leq p \leq \infty$ is $L_{p'}(\Omega)$ with $\frac{1}{p}+\frac{1}{p'} = 1$. The spaces for $0 < p < 1$ have a different structure. Here the constant for the triangle inequality is $2^{1/p}$

$$\|f+g\|_p \leq 2^{1/p}(\|f\|_p + \|g\|_p). \tag{2.5.4}$$

The discrete versions of the L_p spaces are the ℓ_p spaces of sequences $\mathbf{c} = (c_k)_{k\in\mathbb{Z}^d}$ with norm

$$\|c\|_{\ell_p} = \begin{cases} \left(\sum_{k\in\mathbb{Z}^d} |c_k|^p\right)^{1/p}, & 0 < p < \infty \\ \sup_{k\in\mathbb{Z}^d} |c_k|, & p = \infty. \end{cases} \tag{2.5.5}$$

For the L_p spaces there exists a continuous embedding if Ω is a set of finite measure, i.e.

$$L_q(\Omega) \subset L_p(\Omega), \qquad \text{with } \|f\|_p \leq C\|f\|_q \quad \text{for } p \leq q. \tag{2.5.6}$$

The discrete counterpart is

$$\ell_p \subset \ell_q, \qquad \text{with } \left(\sum_{k\in\mathbb{Z}^d} |c_k|^q\right)^{1/q} \leq \left(\sum_{k\in\mathbb{Z}^d} |c_k|^p\right)^{1/p} \qquad \text{for } p \leq q. \tag{2.5.7}$$

An important position among function spaces is taken by smoothness spaces. A natural way to measure the smoothness of a function f is to count the order of differentiability in the classical sense, i.e. to find the maximal index m such that $\frac{\partial^\alpha}{\partial x^\alpha} f$ with $|\alpha| \leq m$ is continuous. The class that is associated to this condition is denoted by $C^m(\Omega)$ for functions f defined on $\Omega \subset \mathbb{R}^d$ and is equipped with the norm

$$\|f\|_{C^m(\Omega)} := \sup_{x\in\Omega} |f(x)| + \sum_{|\alpha|=m} \sup_{x\in\Omega} |\partial^\alpha f(x)|. \tag{2.5.8}$$

Here we measure smoothness of order m in the $L_\infty(\mathbb{R}^d)$ norm. To get more freedom for the choice of the norm, which allows to average the smoothness one may consider SOBOLEV spaces $W^m(L_p(\Omega))$ with $1 \leq p \leq \infty$ and $m = 1, 2, \dots$ which consist of all functions f whose distributional derivatives $\partial^\alpha f$ is in $L_p(\mathbb{R}^d)$. The corresponding norm is

$$\|f\|_{W^m(L_p(\Omega))} := \|f\|_{L_p(\Omega)} + |f|_{W^m(L_p(\Omega))} \tag{2.5.9}$$

$$\text{with} \quad |f|_{W^m(L_p(\Omega))} = \sum_{|\alpha|=m} \|\partial^\alpha f\|_{L_p(\Omega)}. \tag{2.5.10}$$

The mentioned spaces are capable to describe smoothness for integer indexes. As the reader might anticipate it is necessary to extend this notion to fractional smoothness orders.

We begin with fractional smoothness order $0 \leq \alpha \leq 1$. The LIPSCHITZ space $\operatorname{Lip}\alpha$ denotes the set of all functions f on Ω such that for $L > 0$

$$|f(x) - f(y)| \leq L|x - y|^\alpha. \tag{2.5.11}$$

The infimum of all L for which $f \in \operatorname{Lip}\alpha$ is by definition $|f|_{\operatorname{Lip}\alpha}$. In particular, $|f|_{\operatorname{Lip}1} = \|f'\|_{L_\infty}$.

The next approach leads to HÖLDER spaces. If s is a positive real number we put $\{s\} := s - \lfloor s \rfloor$, where $\lfloor s \rfloor$ denotes the largest integer smaller than s. Hölder spaces consist of all functions in $C^{\lfloor s \rfloor}(\Omega)$ with

$$\sum_{|\alpha|=\lfloor s \rfloor} \sup_{x\neq y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x-y|^{\{s\}}} < \infty. \tag{2.5.12}$$

The same idea works for SOBOLEV spaces which gives rise to SLOBODECKIJ spaces $W^s(L_p(\mathbb{R}^d))$ with $s \in \mathbb{R}, s > 0$, with semi-norm

$$\sum_{|\alpha|=\lfloor s \rfloor} \left(\int\int \frac{|D^\alpha f(x) - D^\alpha f(y)|^p}{|x-y|^{d+\{s\}p}} \mathrm{d}x\mathrm{d}y \right)^{1/p}. \tag{2.5.13}$$

One popular way for the generalization of smoothness orders to real numbers is realized by BESSEL potential spaces (or SOBOLEV spaces of fractional order) $H^s(\mathbb{R}^d)$ for $s > 0$ which are defined as containing all functions f with

$$\|f\|^2_{H^s(\mathbb{R}^d)} := \int_{\mathbb{R}^d} (1 + |\omega|^2)^s |\hat{f}(\omega)|^2 \mathrm{d}\omega < \infty. \tag{2.5.14}$$

There exists a norm equivalence between $H^m(\mathbb{R}^d)$ and $W^m(L_2(\Omega))$ for $m \in \mathbb{N}$. A standard book on SOBOLEV spaces is Adams (1978).

In the next section we introduce iterated differences. We use them to build moduli of smoothness which combine properties such as real valued smoothness indices and averaged measure of smoothness. Finally, we see that moduli of smoothness yield to the notion of BESOV spaces which include a wide scale of smoothness spaces and measure real valued smoothness in L_p.

2.5.1 Modulus of Smoothness

We start with iterated differences. For the translation operator

$$\tau_h : f \longmapsto f(\cdot + h), \qquad h \in \mathbb{R}^d \tag{2.5.15}$$

and the identity $\mathcal{I}$ we start with the first difference

$$\Delta_h^1 : \tau_h - \mathcal{I}\,. \tag{2.5.16}$$

By iteration we get the *differences of order* l

$$\Delta_h^l := \Delta_h(\Delta_h^{l-1}), \qquad l = 1, 2, \dots. \tag{2.5.17}$$

With the binomial theorem the iterated difference can be written as follows

$$\Delta_h^l(f, x) = \sum_{k=0}^{l} \binom{l}{k} (-1)^{l-k} f(x + kh), \tag{2.5.18}$$

for all $x \in \Omega_{h,l}$ where

$$\Omega_{h,l} := \{x \in \Omega : \; x + kh \in \Omega, \; k = 1, \dots, l\}, \tag{2.5.19}$$

see e.g. DeVore and Lorentz (1993, Chap.7).

Definition 2.5.1 (Modulus of Smoothness). *The* l-th order L_p-modulus of smoothness *of a function* f *is the function* $\omega_l(f, \cdot, \Omega)_p \;: \mathbb{R}_+ \mapsto \mathbb{R}_+$ *with*

$$\omega_l(f; t, \Omega)_p := \sup_{0 < \|h\|_2 < t} \|\Delta_h^l(f; \cdot)\|_{L_p(\Omega_{h,l})}\;, \qquad 0 \le p \le \infty. \tag{2.5.20}$$

Remark 2.5.1. *If we use the modulus for functions on* $\mathbb{R}^d$ *we write the short notation* $\omega_l(f;t)_p = \omega_l(f;t,\Omega)_p$. *The first order modulus of smoothness* $\omega_1(f;t)_p = \omega(f;t)_p$ *is the so called* modulus of continuity.

Later on we need the following basic properties of the L_p-modulus of smoothness. All proofs are shown in DeVore and Lorentz (1993, Chap.7).

(i) $\omega_r(f;t)_p$ is finite for each t.

(ii) For each f the function $\omega_l(f;t)_p$ is continuous and increasing in t for all p.

(iii) If $f \in L_p(\mathbb{R}^d), 1 \leq p < \infty$ then $\omega_l(f;t)_p \to 0$ for $t \to 0$. If $f \in C_0$ this also holds for $p = \infty$.

(iv) For $1 \leq p \leq \infty$, the triangular inequality shows that

$$\omega_l(f+g;t)_p \leq \omega_l(f;t)_p + \omega_l(g;t)_p, \qquad f,g \in L_p(\mathbb{R}^d). \tag{2.5.21}$$

When $p < 1$ equation (2.5.21) is replaced by

$$\omega_l(f+g;t)_p^p \leq \omega_l(f;t)_p^p + \omega_l(g;t)_p^p. \tag{2.5.22}$$

(v) For any positive integer m and $1 \leq p \leq \infty$

$$\omega_l(f;mt)_p \leq m^l \omega_l(f;t)_p. \tag{2.5.23}$$

For $p < 1$ the corresponding inequality is

$$\omega_l(f;mt)_p^p \leq m^l \omega_l(f;t)_p^p. \tag{2.5.24}$$

(vi) A similar inequality holds for nonintegral factors $\lambda \in \mathbb{R}_+$

$$\omega_l(f;\lambda t)_p \leq (\lambda+1)^l \omega_l(f;t)_p, \qquad \lambda > 0, \tag{2.5.25}$$

or for $p < 1$

$$\omega_l(f;\lambda t)_p^p \leq (\lambda+1)^l \omega_l(f;t)_p^p. \tag{2.5.26}$$

(vii) Since the integral is invariant under translation one has for $h \in \mathbb{R}^d$

$$\omega_l(f(\cdot + h);t)_p = \omega_l(f;t)_p. \tag{2.5.27}$$

(viii) We introduce a short notation for the *dilation* operator.

$$J_n f := |\det M|^{-n/2} f(M^n \cdot), \qquad n \in \mathbb{Z} \tag{2.5.28}$$

Using the identity

$$\|J_n f\|_{L_p(\mathbb{R}^d)} = |\det M|^{n(\frac{1}{2}-\frac{1}{p})} \|f\|_{L_p(\mathbb{R}^d)}, \tag{2.5.29}$$

it follows for the dilation operator

$$\omega_l(J_n f; t)_p \le |\det M|^{n(\frac{1}{2}-\frac{1}{p})} \omega_l(f; \rho(M)^n t)_p. \tag{2.5.30}$$

We recall that $\rho(M) = \max_{i=1,\dots,d} |\lambda_i|$ with λ_i denotes the eigenvalues of the matrix M. Consequently, we will identify $\rho(M)$ with the largest dilation of M.

(ix) The modulus of smoothness is always bounded from above by the Sobolev norm in the following way. For $1 \le p \le \infty$, it is true that

$$\omega_l(f; t)_p \lesssim t^l |f|_{W^{l,p}}, \qquad t \ge 0. \tag{2.5.31}$$

This fact is proved by means of the general Minkowski inequality.

(x) If a function has higher order of smoothness it is natural to measure smoothness of lower order with a modulus of lower order. In particular,

$$\omega_{l+k}(f; t)_p \lesssim t^l \omega_k(f^{(l)}; t)_p, \qquad t \ge 0. \tag{2.5.32}$$

A detailed disquisition on moduli of smoothness and their properties can also be found in Ditzian and Totik (1987).

2.5.2 Besov Spaces

A useful way to create smoothness spaces is to restrict the modulus of smoothness such that we obtain functions that have a common smoothness behaviour. This leads to Besov norms. Here we have three parameters, namely smoothness order α measured in L_p which can be seen as an averaged iterated differences. The third parameter q is secondary and allows to make finer distinctions.

Definition 2.5.2 (Besov seminorm). *Let $\alpha \in \mathbb{R}$, $1 \le p, q \le \infty$ and $l = \lfloor \alpha \rfloor + 1$. The* Besov *seminorm of a function f is defined as*

$$|f|_{\alpha,p,q} := \begin{cases} \left(\int_0^\infty [t^{-\alpha}\omega_l(f,t)_p]^q \frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}} & \text{when } 1 \le q < \infty, \\ \sup_{0<t<\infty} t^{-\alpha}\omega_l(f,t)_p & q = \infty. \end{cases} \tag{2.5.33}$$

For a characterization of BESOV spaces by wavelet expansions we have to deal with sequences of wavelet coefficients. Therefore, we need a discrete version of the BESOV seminorm (2.5.33).

Proposition 2.5.1. *For any α, p, q and $l > \alpha$ as in Definition 2.5.2 we have for a fixed real number $\vartheta > 1$ the relation*

$$|f|_{\alpha,p,q} \sim \left(\sum_{j=0}^{\infty} \vartheta^{\alpha j q} \omega_l(f, \vartheta^{-j})_p^q \right)^{\frac{1}{q}} \tag{2.5.34}$$

Proof. We rewrite the seminorm

$$\begin{aligned} \int_0^\infty [t^{-\alpha}\omega_l(f,t)_p]^q \frac{\mathrm{d}t}{t} &= \int_0^\infty \omega_l(f,t)_p^q \frac{\mathrm{d}t}{t^{\alpha q+1}} \\ &= \sum_{j=0}^{\infty} \int_{\vartheta^{-j}}^{\vartheta^{-j+1}} \omega_l(f,t)_p^q \frac{\mathrm{d}t}{t^{\alpha q+1}}. \end{aligned}$$

Using the fact, that $\omega_l(f, \delta)_p$ is a monotone increasing function and (2.5.25) we majorize the seminorm by

$$\sum_{j=0}^{\infty} (\vartheta - 1)\vartheta^{-j}\omega_l(f, \vartheta^{-j+1})_p^q \vartheta^{j(\alpha q+1)} \leq (\vartheta - 1)(\vartheta + 1)^{rq} \sum_{j=0}^{\infty} \vartheta^{j\alpha q}\omega_l(f, \vartheta^{-j})_p^q$$

and minorize by

$$\sum_{j=0}^{\infty} (\vartheta - 1)\vartheta^{-j}\omega_l(f, \vartheta^{-j})_p^q \vartheta^{(j-1)(\alpha q+1)} = (\vartheta - 1)\vartheta^{-\alpha q-1} \sum_{j=0}^{\infty} \vartheta^{j\alpha q}\omega_l(f, \vartheta^{-j})_p^q.$$

If $q = \infty$ we use the same arguments except that we replace the integrals and the sums by suprema. □

A discrete version of BESOV spaces that is widely used is explained in the following definition.

Definition 2.5.3 (discrete Besov Norm). *The class $B_q^\alpha(L_p(\mathbb{R}^d))$ with $1 \leq p, q \leq \infty$ consists of all functions $f \in L_p(\mathbb{R}^d)$ with $|f|_{B_q^\alpha(L_p(\mathbb{R}^d))} < \infty$, where the seminorm is given by*

$$|f|_{B_q^\alpha(L_p(\mathbb{R}^d))} := \left(\sum_{j=0}^{\infty} (2^{j\alpha}\omega_l(f, 2^{-j})_p)^q \right)^{\frac{1}{q}} \quad \text{with } l = \lfloor \alpha \rfloor + 1. \tag{2.5.35}$$

The discrete BESOV *norm is defined as*

$$\|f\|_{B_q^\alpha(L_p(\mathbb{R}^d))} := \|f\|_{L_p(\mathbb{R}^d)} + |f|_{B_q^\alpha(L_p(\mathbb{R}^d))}. \tag{2.5.36}$$

Remark 2.5.2. *A remarkable fact is that* BESOV *norms are equivalent to many known smoothness norms such as* SOBOLEV, HÖLDER *or* LIPSCHITZ *norms. We briefly summarize some identities. This choice is not complete. A full overview is shown e.g. in Triebel (1983), Triebel (1992) or Runst and Sickel (1996).*

(1.) $C^s = B_\infty^s(L_\infty(\mathbb{R}^d)), \qquad 0 < s, s \in \mathbb{R} \setminus \mathbb{N}$

(2.) $W^s(L_p(\mathbb{R}^d)) = B_p^s(L_p(\mathbb{R}^d)), \qquad 1 \leq p < \infty, 0 < s, s \in \mathbb{R} \setminus \mathbb{N}$

(3.) $H^s = B_2^s(L_2(\mathbb{R}^d)) \qquad s \in \mathbb{R}, s > 0$

The spaces $L_1(\mathbb{R}^d)$, $L_\infty(\mathbb{R}^d)$, $C_0(\mathbb{R}^d)$, $C^m(\mathbb{R}^d)$ *and* $W^m(L_p)$ *for* $m \in \mathbb{N}$ *are not included in the scales* $B_q^s(L_p(\mathbb{R}^d))$ *but the distinctions are very small. For instance, if we replace in the context of* HÖLDER *spaces the first order difference by the 2nd order symmetric difference we obtain the so called* ZYGMUND *spaces* $\mathcal{C}^s$ *which are included in the scale of* BESOV *spaces as* $\mathcal{C}^s = B_\infty^s(L_\infty(\mathbb{R}^d))$ *for real numbers* $0 < s$.

If we increment the secondary index q *for* BESOV *spaces we obtain a larger space, i.e. we have the embedding*

$$B_{q_1}^s(L_p(\mathbb{R}^d)) \subset B_{q_2}^s(L_p(\mathbb{R}^d)), \qquad q_1 < q_2. \tag{2.5.37}$$

This parameter plays a much smaller role than the smoothness index s *since one has for arbitrary* q_1, q_2

$$B_{q_1}^{s_1}(L_p(\mathbb{R}^d)) \subset B_{q_2}^{s_2}(L_p(\mathbb{R}^d)), \qquad s_1 > s_2. \tag{2.5.38}$$

A less trivial embedding is a variant of the so called SOBOLEV *embedding theorem which states for the case of* BESOV *spaces for arbitrary* $s_1, s_2 > 0$ *and* $p_1, p_2 > 1$

$$B_{p_1}^{s_1}(L_{p_1}(\mathbb{R}^d)) \subset B_{p_2}^{s_2}(L_{p_2}(\mathbb{R}^d)), \qquad s_1 - s_2 \geq d(\frac{1}{p_1} - \frac{1}{p_2}). \tag{2.5.39}$$

A visualization of this property is shown in Figure 2.7.

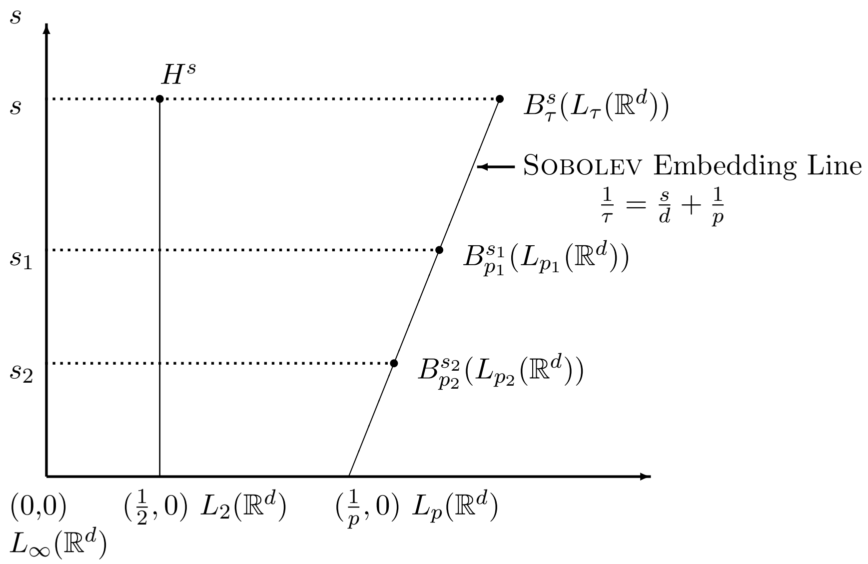


Figure 2.7: Graphical interpretation of the SOBOLEV embedding line.

Chapter 3

Approximation Properties of Shift-Invariant Spaces

In this chapter we go into the world of shift-invariant spaces. In particular we relate polynomial reproduction properties of shift invariant spaces generated by scaling functions to their approximation rate. In particular approximation properties of the level spaces V_j are expressed by the STRANG-FIX conditions. A full description of the connection between approximation and polynomial exactness is specified by the JACKSON or direct estimate which is an important type of inequality that is essential to build the general approximation theory.

With every compactly supported function ϕ we associate the space of all linear combinations of its translates over the lattice points in $\mathbb{Z}^d$,

$$\mathcal{S}(\phi) := \{ \sum_{k \in \mathbb{Z}^d} a_k \phi(\cdot - k) : a_k \in \mathbb{R} \}. \tag{3.0.1}$$

We call this linear space *shift-invariant.*

The approximation properties of shift-invariant spaces are related to polynomial reproduction in $\mathcal{S}(\phi)$. In 1946 SCHOENBERG recognized first that polynomial reproduction could be described by the FOURIER transform of ϕ. Later on, in 1973 STRANG and FIX used FOURIER transforms to describe approximation properties of shift-invariant spaces.

For a good approximation rate the exact reproduction of smooth parts such as polynomials is a necessary property of a system of approximating functions. This is specified by $\Pi_r \subset \mathcal{S}(\phi)$, where Π_r denotes the space of all polynomials of order r (degree $r-1$). Alternatively, this can be described by the condition that for all $q = 0, \ldots, r$ the monomial x^q is expressed according to

$$\sum_{k\in\mathbb{Z}^d} k^q \phi(x-k) = x^q + p_{q-1}(x), \tag{3.0.2}$$

where $x \in \mathbb{R}^d$ and p_{q-1} is a polynomial of degree $q-1$. A proof of this fact is given in Cohen (2003, Th. 2.8.1).

Remark 3.0.3. *Before we continue we recall some notations. In the multivariate setting we often consider d-dimensional multi-indices $k = (k_1, \ldots, k_d)$. We define its length by $|k| := |k_1| + \ldots + |k_d|$. The factorial of k is $k! := k_1! \cdot \ldots \cdot k_d!$. For two multi-indices μ and ν we write $\mu \le \nu$ if $\mu_i \le \nu_i$ for all $i = 1, \ldots, d$. If $\mu \le \nu$ we define the binomial by*

$$\binom{\mu}{\nu} := \frac{\mu!}{\nu!(\mu-\nu)}. \tag{3.0.3}$$

The monomial x^μ for $x = (x_1, \ldots, x_d)$ and $\mu = (\mu_1, \ldots, \mu_d)$ should be understood as $x_1^{\mu_1} \cdot \ldots \cdot x_d^{\mu_d}$.

The partial derivative of a function f with respect to the i-th coordinate is denoted by $\frac{\partial}{\partial x_i}$. For a multi-index $\mu = (\mu_1, \ldots, \mu_d)$ we write D^μ or ∂^μ for the differential operator $\frac{\partial^{\mu_1}}{\partial x_1} \cdots \frac{\partial^{\mu_d}}{\partial x_s}$.

We can also apply a polynomial p to a differential operator D. Then, $p(D)$ denotes the constant coefficient differential operator induced by the polynomial p

For approximation in L_p spaces one considers in general the distance between a given function f and an approximating subset G

$$\mathrm{dist}_p(f, G) = \inf_{g\in G} \|f-g\|_{L_p(\mathbb{R}^d)}. \tag{3.0.4}$$

For the investigation of approximation properties of shift-invariant spaces we confine these considerations. We introduce the space of scaled functions in $\mathcal{S}(\phi)$

$$\mathcal{S}^h := \{g(\cdot/h) : \; g \in \mathcal{S}(\phi) \cap L_p(\mathbb{R}^d), h > 0\}.$$

Given a nonnegative integer α we say that $\mathcal{S}(\phi)$ provides approximation order α if for any function $f \in L_p(\mathbb{R}^d)$

$$\mathrm{dist}_p(f, \mathcal{S}^h) = \mathcal{O}(h^\alpha). \tag{3.0.5}$$

The study of approximation with shift-invariant spaces dates back to the work of de Boor (1987) and was extended in Jia (1995) where the reader also finds a good overview of the literature. The results there state that if ϕ is a compactly supported function on $\mathbb{R}^d$ with $\hat{\phi}(0) \neq 0$ and if $\Pi_{r-1} \subset \mathcal{S}(\phi)$ then $\mathcal{S}(\phi)$ provides approximation order r. It was shown in de Boor (1987) that $\hat{\phi}(0) \neq 0$ and $\Pi_{r-1} \subset \mathcal{S}(\phi)$ together imply that

$$D^\alpha \hat{\phi}(2\pi k) = 0, \qquad \text{for all } |\alpha| < r \text{ and } k \in \mathbb{Z}^d \setminus \{0\}. \tag{3.0.6}$$

Definition 3.0.4 (Strang-Fix Conditions). *If $\hat{\phi}(0) \neq 0$ and condition (3.0.6) holds then we say that ϕ satisfies the* STRANG-FIX condition of order r.

This condition allows to apply a general technique of so-called quasi-interpolants to establish sharp approximation results, see Jia and Lei (1993) for an explicit construction. A survey about the general construction of quasi-interpolants is given in de Boor (1990).

When ϕ satisfies the STRANG-FIX condition of order r it can be shown that $\mathcal{S}(\phi)$ locally contains Π_{r-1}. Cavaretta, Dahmen and Micchelli were the first who proved in Cavaretta, Dahmen, and Micchelli (1991, p.158) polynomial reproducibility of smooth refinable functions for a dyadic multiresolution analysis. Later, Jia improved this result in Jia (1996). He showed that for a $(\mathbf{h}, 2\mathcal{I})$-refinable function $\phi \in W^r(L_1(\mathbb{R}^d))$ and $\hat{\phi}(0) \neq 0$ it follows that $\Pi_r \subset \mathcal{S}(\phi)$. To summarize, in Jia (1995) it was shown that for compactly supported functions $\phi \in \mathbb{R}^d$ with $\hat{\phi}(0) \neq 0$ the following conditions are equivalent for any positive integer r:

1. $\mathcal{S}(\phi)$ provides approximation order r.
2. $\mathcal{S}(\phi)$ contains Π_{r-1}.
3. $D^\alpha \hat{\phi}(2\pi k) = 0$, for all $|\alpha| < r$ and $k \in \mathbb{Z}^d \setminus \{0\}$.

3.1 Generalized Strang-Fix Conditions

In this section we build a bridge between STRANG-FIX conditions for shift-invariant spaces spanned by multivariate translates with isotropic scaling matrices and their approximation properties. Again, here it is not required that the translates necessarily have a tensor product structure.

Jia established STRANG-FIX conditions for translates with isotropic scaling matrices in Jia (1998). The precise result writes as follows:

Theorem 3.1.1 (Jia,1998). *Suppose M is a $d \times d$ isotropic scaling matrix, and $\mathbf{a} = (a_k)_{k\in\mathbb{Z}^d}$ is a finitely supported sequence satisfying*

$$\sum_{k\in\mathbb{Z}^d} a_k = |\det M|.$$

Let ϕ be orthonormalized and $(\mathbf{a}, M)$-refinable. If $\phi \in W^r(L_1(\mathbb{R}^d))$, then one has for all $|\alpha| \leq r$

$$D^\alpha \hat{\phi}(2\pi k) = 0 \qquad \text{for all } k \in \mathbb{Z}^d \setminus \{0\}$$

with a positive $r \in \mathbb{N}$.

In the same work Jia also shows that $\Pi_r \subset \mathcal{S}(\phi)$. In the subsequent propositions we show an alternative proof for this fact. We will see that due to the refineability of the scaling function it constitutes a decomposition of unity. Therefore it follows $\sum_{k\in Z^d} \phi(x-k) \neq 0$ which is assigned to the value of $\hat{\phi}(0)$. Thus we get the first ingredient of the Strang-Fix conditions which is illustrated in the next theorem.

Theorem 3.1.2. *Let $\phi \in L_1(\mathbb{R}^d)$ be a $(\mathbf{a}, M)$-refinable function with $\mathbf{a} \in \ell_1(\mathbb{Z}^d)$. Then*

$$\hat{\phi}(2\pi k) = 0 \quad \text{for all } \ k \in \mathbb{Z}^d \setminus \{0\} \tag{3.1.1}$$

and

$$\hat{\phi}(0) = \sum_{k\in\mathbb{Z}^d} \phi(x-k) \quad a.e.. \tag{3.1.2}$$

Proof. Application of the Fourier transform to the refinement equation (2.1.3) yields

$$\hat{\phi}(\xi) = \frac{1}{|\det M|} a(e^{-iM^{-T}\xi}) \hat{\phi}(M^{-T}\xi), \tag{3.1.3}$$

where we call $a(z)$ the *symbol* associated with the refinement equation (2.1.3). Iteration arises in

$$\hat{\phi}(\xi) = \prod_{j=1}^{n} \left(\frac{1}{|\det M|} a(e^{-i(M^{-T})^j\xi}) \right) \hat{\phi}((M^{-T})^j\xi) \tag{3.1.4}$$

Now we have to distinguish between two cases:

(i) $|a(1)| < |\det M|$
By (3.1.3) follows $\hat{\phi}(0) = \frac{1}{|\det M|} a(1)\hat{\phi}(0)$ and consequently $\hat{\phi}(0) = 0$.

Since $|a(1)| < |\det M|$ we derive $\left|\frac{1}{|\det M|}a(e^{-i(M^{-T})^j\xi})\right| < 1$. Hence, for j sufficiently large we obtain by (3.1.4)

$$\hat{\phi}(\xi) = 0 \quad \text{for all } \xi \in \mathbb{R}^d.$$

So, $\phi = 0$.

(ii) $|a(1)| \geq |\det M|$
Let $\xi = (M^T)^n 2\pi k$ with $k \in \mathbb{Z}^d \setminus \{0\}$ and $n \in \mathbb{N}$. Then we get

$$\begin{aligned}
\hat{\phi}((M^T)^n 2\pi k) &= \prod_{j=1}^{n} \frac{1}{|\det M|}\left(e^{-i(M^{-T})(M^T)^n 2\pi k}\right)\hat{\phi}(2\pi k) \\
&= \left(\frac{1}{|\det M|}a(1)\right)^n \hat{\phi}(2\pi k).
\end{aligned}$$

Thus, it appears that

$$|\hat{\phi}(2\pi k)| \leq |\hat{\phi}((M^T)^n 2\pi k)| \leq \lim_{n\to\infty} |\hat{\phi}((M^T)^n 2\pi k)| = 0,$$

where the last equation follows by the RIEMANN-LEBESGUE theorem.

Because $\phi \in L_1(\mathbb{R}^d)$ it follows $\sum_{k\in\mathbb{Z}^d} \phi(\cdot - k) \in L_1([0,1)^d)$ and we can expand $\sum_{k\in\mathbb{Z}^d} \phi(\cdot - k)$ in a FOURIER series. This implies

$$\sum_{k\in\mathbb{Z}^d} \phi(x-k) \sim \sum_{l\in\mathbb{Z}^d} c_l e^{2\pi i l x}$$

with

$$\begin{aligned}
c_l &= \int_{[0,1)^d} \left(\sum_{k\in\mathbb{Z}^d} \phi(x-k)\right) e^{-i2\pi l x}\mathrm{d}x \\
&= \int_{\mathbb{R}^d} \phi(x) e^{-i2\pi l x}\mathrm{d}x = \hat{\phi}(2\pi l).
\end{aligned}$$

In Stein and Weiss (1971, p.249) it is shown that for every $f \in L_1(\mathbb{T}^d)$ with $\mathbb{T}^d = \mathbb{R}^d/2\pi\mathbb{Z}^d$ and $(a_k)_{k\in\mathbb{Z}^d}$ with $\sum_{k\in\mathbb{Z}^d} |a_k| < \infty$ one has an expansion of the form

$$f(x) = \sum_{k\in\mathbb{Z}^d} a_k e^{ikx} \quad a.e.\,.$$

Equation (3.1.1) puts all FOURIER coefficients to zero except for $l = 0$. Using this fact we obtain the desired second equation (3.1.2).

□

The same result as equation (3.1.2) can be obtained without refineability of the generator functions with the additional assumptions that the STRANG-FIX conditions (3.1.1) are valid and ϕ is in $C_0(\mathbb{R}^d)$.

Lemma 3.1.3. *Let $\phi \in C_0(\mathbb{R}^d)$ be a $(\mathbf{a}, M)$-refinable function with $\mathbf{a} \in \ell_1(\mathbb{Z}^d)$ and $\hat{\phi}(2\pi k) = 0$ for all $k \in \mathbb{Z}^d \setminus \{0\}$. Then*

$$\hat{\phi}(0) = \sum_{k \in \mathbb{Z}^d} \phi(x-k), \quad a.e. \tag{3.1.5}$$

Proof. Since $\phi \in C_0(\mathbb{R}^d)$ the sum $\sum_{k \in \mathbb{Z}^d} \phi(k)$ is in $L_1([0,1)^d)$. Analogously, we proceed as in the previous proof and get the assertion. □

Corollary 3.1.4. *Let $\phi \in \mathcal{L}_p$ be a $(\mathbf{a}, M)$-refinable ℓ_p-stable function with $\mathbf{a} \in \ell_1(\mathbb{Z}^d)$. Then*

$$\sum_{k \in \mathbb{Z}^d} \phi(x-k) = 1 \quad a.e. \ . \tag{3.1.6}$$

Proof. We apply Theorem 3.1.2, more precisely equation (3.1.2). It suffices to show that $\hat{\phi}(0) \neq 0$. Then we can divide ϕ by this constant and get the desired result.
Since ϕ is ℓ_p-stable we have $\sup_{k \in \mathbb{Z}^d} \left|\hat{\phi}(\xi + 2\pi k)\right| > 0$. With $\xi = 0$ and (3.1.1) we conclude $\sup_{k \in \mathbb{Z}^d} \left|\hat{\phi}(2\pi k)\right| = \hat{\phi}(0)$. □

3.2 Reproduction of Polynomials

Now we incorporate STRANG-FIX conditions. In particular we show that STRANG-FIX conditions of order r for refinable functions imply the reproduction of polynomials up to order r.

First we extend the definition of the semi-discrete convolution for a function and a sequence to the semi-discrete convolution of two functions. Let ϕ and $f \in \mathcal{L}_p$. We define

$$\phi *' f := \phi *' (f|_{\mathbb{Z}^d})(x) = \sum_{k \in \mathbb{Z}^d} f(k)\phi(\cdot - k). \tag{3.2.1}$$

Now we consider the following class of polynomials associated with a refinable function ϕ:

$$\Pi_\phi := \left\{p \in \Pi : \ \phi *' p \in \Pi\right\}. \tag{3.2.2}$$

A surprising fact is that this class are those polynomials which are also contained in $\mathcal{S}(\phi)$.

Theorem 3.2.1. *Let* $\sum_{k\in\mathbb{Z}^d}\phi(k)=1$. *Then*

$$\Pi\cap\mathcal{S}(\phi)=\Pi_\phi. \tag{3.2.3}$$

Proof. We show that both sets are mutually included in each other.

Let $p\in\Pi_\phi$. We see that $\phi *' p(k) = p*'\phi(k)$ for all integer points $k\in\mathbb{Z}^d$. If we are given the values of $p*'\phi$ at all points $k\in\mathbb{Z}^d$ the polynomial $p*'\phi$ is uniquely determined and so is $\phi *' p$. Thus, we get

$$(\phi *' p)(x) = (p*'\phi)(x), \quad x\in\mathbb{R}^d. \tag{3.2.4}$$

This means we can also characterize the space Π_ϕ by

$$\Pi_\phi := \left\{p\in\Pi:\ \phi *' p = p *' \phi\right\}. \tag{3.2.5}$$

We show for all $f\in\mathcal{S}(\phi)$

$$\phi *' f = f *' \phi. \tag{3.2.6}$$

The left term of equation (3.2.6) simplifies to

$$\begin{aligned}\phi *' f &= \phi *' (\phi *' c) = \sum_{l\in\mathbb{Z}^d}(\phi *' c)(l)\phi(\cdot - l) && (3.2.7)\\ &= \sum_{l\in\mathbb{Z}^d}\left(\sum_{k\in\mathbb{Z}^d}c_k\phi(l-k)\right)\phi(\cdot - l) && (3.2.8)\\ &= \sum_{k\in\mathbb{Z}^d}c_k\sum_{l\in\mathbb{Z}^d}\phi(l-k)\phi(\cdot - l). && (3.2.9)\end{aligned}$$

For the right term of equation (3.2.7) we arrive at

$$\begin{aligned}f *' \phi &= \sum_{l\in\mathbb{Z}^d}\phi(l)f(\cdot - l) = \sum_{l\in\mathbb{Z}^d}\phi(l)(\phi *' c)(\cdot - l)\\ &= \sum_{l\in\mathbb{Z}^d}\phi(l)(\sum_{k\in\mathbb{Z}^d}c_k\phi(\cdot - l - k)\\ &= \sum_{k\in\mathbb{Z}^d}c_k\sum_{l\in\mathbb{Z}^d}\phi(l)\phi(\cdot - l - k)\,, \quad \tilde{l} = l + k\\ &= \sum_{k\in\mathbb{Z}^d}c_k\sum_{\tilde{l}\in\mathbb{Z}^d}\phi(\tilde{l}-k)\phi(\cdot - \tilde{l}),\end{aligned}$$

which concludes the equality. By combination of (3.2.5) and (3.2.6) we obtain the left inclusion $\Pi\cap\mathcal{S}(\phi)\subset\Pi_\phi$.

Now we show the other direction of equation (3.2.3), that is the right inclusion $\Pi \cap \mathcal{S}(\phi) \supset \Pi_\phi$. We consider the mapping

$$L_\phi : \Pi_\phi \longrightarrow \Pi,\ p \longmapsto \phi *' p$$

Since $\dim \Pi_\phi < \infty$ it remains to show that L_ϕ is injective. It follows that

$$\begin{aligned} \phi *' p &= p *' \phi = \sum_{k\in\mathbb{Z}^d} p(\cdot - k)\phi(k) \\ &= p(\cdot)\sum_{k\in\mathbb{Z}^d} \phi(k) - \sum_{k\in\mathbb{Z}^d} (p(\cdot) - p(\cdot - k))\phi(k) \\ &= p - \tilde{p}, \end{aligned}$$

where $\tilde{p}$ is a polynomial in Π_n with degree $n < p$. □

Following the line of Dahmen and Micchelli (1983) we introduce the class

$$\mathcal{P}_\phi := \left\{ p \in \Pi :\ (p(D)\hat{\phi})(2\pi k) = 0,\ k \in \mathbb{Z}^d \setminus \{0\} \right\}, \tag{3.2.10}$$

where $p(D)$ denotes the constant coefficient differential operator induced by the polynomial p. It is necessary to restrict this class. To this end we work with affine invariant subspaces of $\mathcal{P}_\phi$.

Definition 3.2.1 (Affine Invariant Subspaces). *A set $\mathcal{P}$ is said to be* affine invariant *if*

(i) $\mathcal{P}$ is a subspace of $\mathcal{P}_\phi$,

(ii) whenever $p \in \mathcal{P}$ also $p(ax + y) \in \mathcal{P}$ for $a \in \mathbb{C}$, $y \in \mathbb{R}^d$.

We point out the connection between locality of a function and smoothness of its Fourier transform. To be more accurate, there exists an amplification of this relation. The „most localized“ functions – the functions with compact support have the „smoothest“, thus analytical Fourier transforms. This is the central proposition of the well known Paley-Wiener Theorem. To formulate it we need the following definition.

Definition 3.2.2 (Exponential Type). *An entire function G is of* exponential type σ *if for all positive real numbers $\sigma, \epsilon > 0$ there exists a positive constant A_ϵ such that*

$$|G(z)| \leq A_\epsilon e^{(\sigma+\epsilon)\|z\|_2} \qquad z \in \mathbb{C}^d. \tag{3.2.11}$$

The PALEY-WIENER Theorem states the following:

Theorem 3.2.2 (Paley-Wiener). *A function $g \in L_2(\mathbb{R}^d)$ is the* FOURIER *transform of a function f with* $\operatorname{supp} f \subset]-\frac{\sigma}{2\pi}, \frac{\sigma}{2\pi}]$ *if and only if it is a restriction of an entire function G of exponential type σ on the real line.*

A proof of this theorem is shown e.g. in Stein and Weiss (1971, Chap. III, Sec. 4).

Theorem 3.2.2 states that $\hat{\phi}$ is an entire function of exponential type if $\phi \in C_0(\mathbb{R}^d)$. If furthermore $\hat{\phi}(0) \neq 0$ we get that $\frac{1}{\hat{\phi}(0)}$ is an analytical function. That is, we can expand it into a power series in some neighborhood of the origin

$$(\hat{\phi}(\xi))^{-1} = \sum_{k \geq 0} c_k \xi^k \tag{3.2.12}$$

with suitable coefficients $c_k \in \mathbb{R}$. This motivates the definition of the following mapping for a smooth function $f \in C_0^\infty(\mathbb{R}^d)$

$$(Lf)(x) := \sum_{k \in \mathbb{Z}^d} c_k (-i)^{|k|} \partial^k f(x) \tag{3.2.13}$$

which follows the idea of TAYLOR series.

Lemma 3.2.3. *Let $\phi \in C_0(\mathbb{R}^d)$ and $\hat{\phi}(0) \neq 0$. Further, let $\mathcal{P}$ be an affine invariant subspace of $\mathcal{P}_\phi$. Then*

$$Lp \in \mathcal{P} \quad \textit{for all}\, p \in \mathcal{P}. \tag{3.2.14}$$

Proof. We split the proof in three parts. It is sufficient to show the following three assertions:

(i) $\partial^y p \in \mathcal{P}$ for every $y \in \mathbb{Z}^d$.

(ii) $1 \in \mathcal{P}$.

(iii) whenever $p = q_1 + \ldots + q_n$ where each q_i is a homogeneous polynomial with mutual distinct orders, then $q_i \in \mathcal{P}$, $i = 1, \ldots, n$.

To prove (i) it is sufficient to show $x^\alpha \in \mathcal{P} \implies \partial^{e_j} x^\alpha \in \mathcal{P}$ for any $\alpha \in \mathbb{Z}^d$. (Here e_j denotes the unit vector with entry 1 at position j) Linear combinations of the monomials concludes the assertion. It follows

$$\begin{aligned} \partial^{e_j} x^\alpha &= \partial^{e_j} (x_1^{\alpha_1} \cdot \ldots \cdot x_d^{\alpha_d}) \\ &= x_1^{\alpha_1} \cdot \ldots \cdot x_{j-1}^{\alpha_{j-1}} (\alpha_j x_j^{\alpha_j - 1}) x_{j+1}^{\alpha_{j+1}} \cdot \ldots \cdot x_d^{\alpha_d}. \end{aligned}$$

Since $\mathcal{P}$ is affine invariant we also know that $(ax+y)^\alpha \in \mathcal{P}$. In particular this is true for $a = 1$ and $y = e_j$.

$$\begin{aligned}(x+e_j)^\alpha &= x_1^{\alpha_1}\cdot\ldots\cdot x_{j-1}^{\alpha_{j-1}}(x_j+1)^{\alpha_j}x_{j+1}^{\alpha_{j+1}}\cdot\ldots\cdot x_d^{\alpha_d}\\ &= x_1^{\alpha_1}\cdot\ldots\cdot x_{j-1}^{\alpha_{j-1}}(\sum_{k=0}^{\alpha_j}\binom{\alpha_j}{k}x_j^k)x_{j+1}^{\alpha_{j+1}}\cdot\ldots\cdot x_d^{\alpha_d}\\ &= \sum_{k=0}^{\alpha_j} q_k(x).\end{aligned}$$

Each q_k is a homogeneous polynomial with distinct order. Hence, we can apply (iii) and conclude that $\partial^{e_j}x^\alpha \in \mathcal{P}$. All other cases for $a \neq 1$ and arbitrary y follow by linear combinations of the result above.

From fact (i) we know that all derivatives of p are contained in $\mathcal{P}$ and hence also constant functions. Thus it follows fact (ii).

A proof of fact (iii) can be found in Dahmen and Micchelli (1983). □

A minimal requirement for a local approximation operator is that it reproduces constant functions. This carries over in the setting with the biorthogonal pair of scaling functions $(\phi, \tilde{\phi})$ associated with the projector P_j as defined in (2.2.1) in the following way

$$P_j 1 = \sum_k \langle 1, \tilde{\phi}_{j,k}\rangle \phi_{j,k} = 1. \qquad (3.2.15)$$

This ensures that in particular constant functions are generated by the translates of ϕ. Those general approximation operators, when the function $\tilde{\phi}_{j,k}$ is not necessarily dual to ϕ, are called *quasi interpolant*. These interpolants were introduced in de Boor and Fix (1973) for spline approximation.

Theorem 3.2.4 (Reproduction of Polynomials). *Let $\phi \in C_0(\mathbb{R}^d)$ and $\hat{\phi}(0) \neq 0$. Further, let $\mathcal{P}$ be an affine invariant subspace of $\mathcal{P}_\phi$. Then, the quasi interpolant $T : f \longmapsto Tf$ induced by the mapping*

$$(Tf)(x) := \sum_{k\in\mathbb{Z}^d} (Lf)(k)\phi(x-k) = \phi *' (Lf)(x) \qquad (3.2.16)$$

reproduces all polynomials $p \in \mathcal{P}$, that is

$$Tp = p \quad \textit{for all } p \in \mathcal{P}.$$

Proof. Let $p \in \mathcal{P}$. We define

$$\tilde{\varphi}(y) := (Lp)(y)\phi(x-y). \tag{3.2.17}$$

Note that this function is not necessarily refinable but it is in $C_0(\mathbb{R}^d)$. Then we derive

$$\begin{aligned}
\hat{\tilde{\varphi}}(\xi) &= \int_{\mathbb{R}^d} (Lp)(y)\phi(x-y)e^{-iy\xi}\mathrm{d}y \\
&= -\int_{\mathbb{R}^d} (Lp)(x-z)\phi(z)e^{-i(x-z)\xi}\mathrm{d}z \\
&= -e^{-ix\xi}\int_{\mathbb{R}^d} (Lp)(x-y)\phi(z)e^{iz\xi}\mathrm{d}z.
\end{aligned}$$

Let $p(x) = x^\alpha$, $\alpha \in \mathbb{Z}^d$. We show

$$(p(D+x))\,\hat{\phi}(\xi) = \int_{\mathbb{R}^d} p(x-z)\phi(z)e^{-iz\xi}\mathrm{d}z. \tag{3.2.18}$$

For this we use induction by $|\alpha|$. The statement is true for $|\alpha| = 0$. Assume that it is also true for $|\alpha| = n$, $|\beta| = n+1$. Note that then $\beta = \alpha + e_j$ with the unit vector $e_j = (0,\dots,0,1,0,\dots,0)$ with entry 1 on position j. We derive

$$\begin{aligned}
(D+x)^{\alpha+e_j}\hat{\phi}(\xi) &= (D+x)^{e_j}(D+x)^\alpha\hat{\phi}(\xi) \\
&= (D+x)^{e_j}\int_{\mathbb{R}^d}(x-z)^\alpha\phi(z)e^{-iz\xi}\mathrm{d}z \\
&= \int_{\mathbb{R}^d}(x-z)^\alpha\left(\phi(z)e^{-iz\xi}x_j - \phi(z)(\frac{1}{i}\frac{\partial}{\partial\xi_j}e^{-iz\xi})\right)\mathrm{d}z \\
&= \int_{\mathbb{R}^d}(x-z)^\alpha\phi(z)e^{-iz\xi}(x_j-z_j)\mathrm{d}z \\
&= \int_{\mathbb{R}^d}(x-z)^\beta\phi(z)e^{-iz\xi}\mathrm{d}z,
\end{aligned}$$

which proves equation (3.2.18). Employing from (3.2.14) that $Lp \in \mathcal{P}$ we expand Lp as

$$(Lp)(x) = \sum_{|\alpha|=0}^{N} a_\alpha x^\alpha$$

and obtain

$$\begin{aligned}\hat{\tilde{\varphi}}(\xi) &= e^{-ix\xi}\int_{\mathbb{R}^d}\sum_{|\alpha|=0}^{N} a_\alpha (x-z)^\alpha \phi(z) e^{iz\xi}\mathrm{d}z \\ &= e^{-ix\xi}\sum_{|\alpha|=0}^{N} a_\alpha \int_{\mathbb{R}^d}(x-z)^\alpha \phi(z) e^{iz\xi}\mathrm{d}z \\ &= e^{-ix\xi}\sum_{|\alpha|=0}^{N} a_\alpha (D+x)^\alpha \hat{\phi}(-\xi) \\ &= e^{-ix\xi}(Lp)(D+x)\hat{\phi}(-\xi).\end{aligned}$$

This yields the assertion

$$\hat{\tilde{\varphi}}(\xi) = e^{-ix\xi}((Lp)(-iD+x)\hat{\phi})(-\xi). \tag{3.2.19}$$

Since $p \in \mathcal{P}$ we get by Lemma 3.2.3 that $Lp \in \mathcal{P}$. Because $\mathcal{P}$ is affine invariant we also have that $(Lp)(D+x) \in \mathcal{P}$. This means for all $k \in \mathbb{Z}^d \setminus \{0\}$

$$(Lp)(D+x)\hat{\phi}(2\pi k) = 0$$

and hence

$$\hat{\tilde{\varphi}}(2\pi k) = 0, \qquad \text{for all } k \in \mathbb{Z}^d \setminus \{0\}.$$

We can use this assertion together with the fact that $\hat{\tilde{\varphi}} \in C_0(\mathbb{R}^d)$ for Theorem 3.1.2. We get

$$\begin{aligned}(Tp)(x) &= \sum_{k\in\mathbb{Z}^d}(Lp)(k)\phi(x-k) = \sum_{k\in\mathbb{Z}^d}\tilde{\varphi}(k) \\ &= \hat{\tilde{\varphi}}(0) = (Lp)(-iD+x)\hat{\phi}(0) \\ &= \left(\sum_{k\in\mathbb{Z}^d} c_k(-i)^{|k|}\partial^k p(D+x)\hat{\phi}\right)(0).\end{aligned} \tag{3.2.20}$$

If we assume the representation $p(x) = \sum_{|l|\leq r} b_l x^l$ for a polynomial $p \in \Pi_r$ then it follows for its derivative

$$\partial^k p(x) = \sum_{\substack{l \\ |k|\leq|l|\leq r}} b_l \frac{l!}{(l-k)!} x^{l-k}. \tag{3.2.21}$$

Inserting (3.2.21) in (3.2.20) yields

$$
\begin{aligned}
(Tp)(x) &= \left(\sum_{|k|\le r} c_k (-i)^{|k|} \sum_{\substack{l \\ |k|\le|l|\le r}} b_l \frac{l!}{(l-k)!} (D+x)^{l-k}\hat{\phi} \right)(0) \\
&= \left(\sum_{|k|\le r} c_k (-i)^{|k|} \sum_{\substack{l \\ |k|\le|l|\le r}} b_l \frac{l!}{(l-k)!} \right. \\
&\qquad \left. \cdot \sum_{\substack{\gamma \\ 0\le|\gamma|\le|l-k|}} \binom{l-k}{\gamma} D^\gamma x^{l-k-\gamma}\hat{\phi} \right)(0).
\end{aligned}
$$

If we use that $\binom{l-k}{\gamma} = 0$ for $|\gamma| > |l-k|$ we continue by

$$
\begin{aligned}
(Tp)(x) &= \left(\sum_{|k|\le r} c_k (-i)^{|k|} \sum_{|l|=|k|}^{r} b_l \frac{l!}{(l-k)!} \right. \\
&\quad \left. \cdot \sum_{|\gamma|=0}^{|k|} \frac{(l-k)!}{(l-k-\gamma)!\gamma!} D^\gamma x^{l-k-\gamma}\hat{\phi} \right)(0) \\
&= \left(\sum_{|k|\le r} c_k (-i)^{|k|} \sum_{|\gamma|=0}^{|k|} \frac{1}{\gamma!} D^\gamma \sum_{|l|=|\gamma+k|}^{r} b_l \frac{l!}{(l-k)!} x^{l-k-\gamma}\hat{\phi} \right)(0) \\
&= \left(\sum_{|k|\le r} c_k (-i)^{|k|} \sum_{|\gamma|=0}^{|k|} \frac{1}{\gamma!} D^\gamma \partial^{k+\gamma} p(x)\hat{\phi} \right)(0) \\
&= \sum_{|k|\le r} \sum_{|\gamma|\le r} c_k (-i)^{|\gamma+k|} \frac{1}{\gamma!} \partial^\gamma \hat{\phi}(0) \partial^{\gamma+k} p(x) \\
&= \sum_{|l|=0}^{2r} (-i)^{|l|} \partial^l p(x) \left(\sum_{\substack{k;\gamma \\ k+\gamma=l}} c_k \frac{1}{\gamma!} \partial^\gamma \hat{\phi}(0) \right).
\end{aligned}
$$

We are done if we show

$$
\sum_{\substack{k;\gamma \\ k+\gamma=l}} c_k \frac{1}{\gamma!} \partial^\gamma \hat{\phi}(0) = \delta_{0l}. \tag{3.2.22}
$$

We know by (3.2.12)

$$1 = \left(\sum_{k \geq 0} c_k \xi^k \right) \hat{\phi}(\xi). \tag{3.2.23}$$

If $l = 0$ so are $k = 0$ and $\gamma = 0$ and we get $c_0 \hat{\phi}(0) = 0$. Now, let $l \neq 0$. Differentiation of order l in (3.2.12) results in

$$\begin{aligned} 0 &= \partial^l \left(\sum_{k \geq 0} c_k \xi^k \right) \hat{\phi}(\xi) \\ &= \sum_{0 \leq \mu \leq l} \binom{l}{\mu} \partial^\mu \left(\sum_{k \geq 0} c_k \xi^k \right) \partial^{l-\mu} \hat{\phi}(\xi) \\ &= \sum_{0 \leq \mu \leq l} \binom{l}{\mu} \sum_{k \geq \mu} c_k \frac{k!}{(k-\mu)!} \xi^{k-\mu} \partial^{l-\mu} \hat{\phi}(\xi). \end{aligned}$$

For $\xi = 0$ we obtain

$$\begin{aligned} 0 &= \sum_{0 \leq \mu \leq l} \binom{l}{\mu} c_\mu \mu! \hat{\phi}(0) \\ &= \sum_{0 \leq \mu \leq l} \frac{l!}{(l-\mu)!} c_\mu \mu! \hat{\phi}(0) \\ &= \sum_{0 \leq \gamma \leq l} c_{l-\gamma} \frac{1}{\gamma!} \partial^\gamma \hat{\phi}(0) \\ &= \sum_{\substack{k;\gamma \\ 0 \leq k, \gamma \leq l \\ k+\gamma = l}} c_k \frac{1}{\gamma!} \partial^\gamma \hat{\phi}(0), \end{aligned}$$

where we used the substitution $\gamma = l - \mu$. Thus, we have shown (3.2.22) which implies $Tp = p$. □

The central theorem of this section states that Strang-Fix conditions ensure that shift-invariant functions from $\mathcal{S}(\phi)$ reproduce polynomials. The precise formulation is stated in the following corollary.

Corollary 3.2.5. *Let $\phi \in C_0(\mathbb{R}^d)$ and $\hat{\phi}(0) \neq 0$. If ϕ satisfies the generalized* Strang-Fix *conditions of order r, i.e.*

$$(\partial^\alpha \hat{\phi})(2\pi k) = 0, \quad \text{for all } k \in \mathbb{Z}^d \setminus \{0\},\ |\alpha| \leq r, \tag{3.2.24}$$

then

$$\Pi_r \subset \mathcal{S}(\phi). \tag{3.2.25}$$

Proof. The STRANG-FIX conditions (3.2.24) ensure that $\Pi_r \subset \mathcal{P}_\phi$. Since Π_r is affine invariant we obtain by application of Theorem 3.2.4 that $Tp = p$ for all $p \in \Pi_r$. The representation

$$p(x) = (Tp)(x) = \sum_{k\in\mathbb{Z}^d} (Lp)(k)\phi(x-k)$$

shows us that p is written in terms of integer translates of the function ϕ which proves that p is an element of $\mathcal{S}(\phi)$. □

Remark 3.2.1. *Since Π_r is invariant of the change of scale it is obvious that all polynomials in Π_r are also contained in the spaces V_j of the multiresolution analysis. In Cohen (2003, Th. 2.8.1) it is shown that condition (3.0.2) is equivalent to the fact that for all $q = 0,\ldots,r$ the polynomial x^q can be expanded as*

$$x^q = \sum_{k\in\mathbb{Z}^d} (k^q + p_{q-1}(k))\phi(x-k), \tag{3.2.26}$$

with a polynomial p_{q-1} of degree $q-1$. If we consider a now biorthogonal pair of scaling functions $(\phi,\tilde{\phi})$ in the sense of (2.1.19) we deduce that for all $q = 1,\ldots,r-1$ it follows

$$k^q + p_{q-1}(k) = \int_{\mathbb{R}^d} x^q \tilde{\phi}(x-k)\mathrm{d}x. \tag{3.2.27}$$

This is very close to the vanishing moment condition. Also we conclude that $P_j p = p$ for all polynomials $p \in \Pi_{r-1}$.

In this section we have seen that non-separable scaling functions with isotropic scaling matrices satisfy STRANG-FIX conditions of a certain order. STRANG-FIX conditions imply polynomial reproduction and polynomial reproduction implies a certain approximation order. In the next section relate the polynomial reproduction to particular error estimates of the type

$$\inf_{g\in V_j} \|f-g\|_{L_p(\mathbb{R}^d)} \lesssim \|M^{-j}\|_\infty^{r+1} |f|_{W^{r+1}(L_p(\mathbb{R}^d))} \tag{3.2.28}$$

for functions in the SOBOLEV space $W^{r+1}(L_p(\mathbb{R}^d))$. It turns out that STRANG-FIX conditions exactly determine the order of approximation in these spaces.

3.3 Jackson Type Estimates

In the following it is necessary to extend the linear functional

$$L : \mathcal{P} \longrightarrow \mathbb{R},\ p \longmapsto (Lp)(0)$$

with L defined in (3.2.13) to a linear functional $\tilde{L}$ on $L_\infty([-1,1]^d)$. Then we ensure that translation is an isometry and restriction to closed subsets in $\mathbb{R}^d$ is a contraction. The HAHN-BANACH Theorem states that there exists such an extension with

$$\begin{aligned} |\tilde{L}(f)| &\leq \|\tilde{L}\| \cdot \|f\|_{L_\infty([-1,1]^d)}, \quad \text{for all } f \in L_\infty([-1,1]^d) && (3.3.1)\\ \tilde{L}(p) &= (Lp)(0), \quad \text{for all } p \in \mathcal{P}. && (3.3.2)\end{aligned}$$

We set

$$(T_M f)(x) := \sum_{k\in\mathbb{Z}^d} \tilde{L}(f(M(\cdot + k)))\phi(M^{-1}x - k) \tag{3.3.3}$$

with a $\mathbb{Z}^{d\times d}$ scaling matrix M.

Lemma 3.3.1. *Let $\phi \in C_0(\mathbb{R}^d)$ and $\mathcal{P}$ be an affine invariant subspace of $\mathcal{P}_\phi$. Then, there exists a positive constant $c \in \mathbb{R}$ such that*

$$|f(x) - (T_M f)(x)| \lesssim \inf_{p\in\mathcal{P}} \|(f-p)(M \cdot + x)\|_{L_\infty([-1,1]^d c)} \tag{3.3.4}$$

Proof. First we show that there exists a positive constant c independent of M such that

$$|(T_M f)(x)| \lesssim \|f(M \cdot + x)\|_{L_\infty([-1,1]^d c)}. \tag{3.3.5}$$

For that we write

$$\begin{aligned} |(T_M f)(x)| &\lesssim \sum_{k\in\mathbb{Z}^d} |\tilde{L}(f(M(\cdot + k)))| \cdot |\phi(M^{-1}x - k)| \\ &\lesssim \sum_{k\in\mathbb{Z}^d} \|f(M(\cdot + k))\|_{L_\infty([-1,1]^d)} |\phi(M^{-1}x - k)| \\ &\lesssim \sum_{k\in\mathbb{Z}^d} \|f(M(\cdot + x))\|_{L_\infty([-1,1]^d - M^{-1}x + k)} |\phi(M^{-1}x - k)|. \end{aligned}$$

Since ϕ has compact support it follows assertion (3.3.5).
Now, for all $p \in \mathcal{P}$ it is true that

$$\begin{aligned}(T_M p)(x) &= \sum_{k\in\mathbb{Z}^d} \tilde{L}(p(M(\cdot+k)))\phi(M^{-1}x-k)\\ &= \sum_{k\in\mathbb{Z}^d} (Lp(M(\cdot+k)))|_0\phi(M^{-1}x-k)\\ &= \sum_{k\in\mathbb{Z}^d} (Lp(Mk))\phi(M^{-1}x-k)\\ &= (Tp(M\cdot))(M^{-1}x) = p(M\cdot)(M^{-1}x) = p(x).\end{aligned}$$

Then we go on with

$$\begin{aligned}|f(x)-T_M f(x)| &\lesssim |f(x)-p(x)| + |T_M f(x) - T_M p(x)|\\ &\lesssim |f(x)-p(x)| + |T_M(f-p)(x)|\\ &\lesssim |f(x)-p(x)| + \|(f-p)(M\cdot+x)\|_{L_\infty([-1,1]^d c)},\end{aligned}$$

where the last inequality follows by (3.3.5). Combining this and

$$\begin{aligned}|f(x)-p(x)| &= |(f(M\cdot+x)-p(M\cdot+x))|_0|\\ &\lesssim \|f(M\cdot+x)-p(M\cdot+x)\|_{L_\infty([-1,1]^d c)}\end{aligned}$$

we obtain

$$|f(x)-T_M f(x)| \lesssim \inf_{p\in\mathcal{P}} \|(f-p)(M\cdot+x)\|_{L_\infty([-1,1]^d c)}.$$

□

Lemma 3.3.2. *Let $\phi \in C_0(\mathbb{R}^d)$ and $\Pi_r \subset \mathcal{P}_\phi$. Furthermore, let $f \in W^r(L_p(\mathbb{R}^d))$. Then we have*

$$|f(x)-T_M f(x)| \lesssim \|M\|_\infty^{r-\frac{d}{p}} |f|_{W^r(L_p(\|M\|_\infty c([-1,1]^d)+x))}. \tag{3.3.6}$$

Proof. We use the so called Whitney type estimates to show the proposition. They state the following:

Let I be a cube in $\mathbb{R}^d$ and $f \in W^r(L_p(I))$. Then one has for each $\mu = \frac{r}{d} - \frac{1}{p} + \frac{1}{\gamma}$, $1 \le \gamma \le \infty$

$$E_r(f,I)_\gamma := \inf_{p\in\Pi_r} \|f-p\|_{L_\gamma(I)} \lesssim |I|^\mu |f|_{W^r(L_p(I))}. \tag{3.3.7}$$

Brudnyi (1970) was the first who showed that this inequality holds for $1 \leq p \leq \infty$. For other works concerning WHITNEY type estimates the reader is referred to the standard finite element literature, e.g. Deny and Lions (1954) or Ciarlet (1991).

From equation (3.3.4) we deduce with (3.3.7)

$$\begin{aligned} |f(x) - (T_M f)(x)| &\lesssim \inf_{p \in \mathcal{P}} \|(f-p)(M \cdot + x)\|_{L_\infty([-1,1]^d c)} \\ &\lesssim \inf_{p \in \Pi_r} \|(f-p)(\cdot)\|_{L_\infty(\|M\|_\infty c[-1,1]^d + x)} \\ &\lesssim \|M\|_\infty^{d(\frac{r}{d} - \frac{1}{p})} |f|_{W^r(L_p(\|M\|_\infty c([-1,1]^d)+x))} \\ &\lesssim \|M\|_\infty^{r - \frac{d}{p}} |f|_{W^r(L_p(\|M\|_\infty c([-1,1]^d)+x))}, \end{aligned}$$

which shows equation (3.3.6). □

Now we arrive at the main theorem of this chapter. The full relation between smoothness of a function and the corresponding approximation rate by shift-invariant spaces is written in the following theorem.

Theorem 3.3.3 (Jackson Type Estimate). *Let $f \in W^{r+1}(L_p(\mathbb{R}^d))$. Further, let $\phi \in C_0(\mathbb{R}^d)$, $\phi \in W^r(L_1(\mathbb{R}^d))$ be a $(\mathbf{a}, M)$-refinable ℓ_2-stable function. If ϕ satisfies the generalized* STRANG-FIX *conditions of order r, i.e.*

$$(\partial^\alpha \hat{\phi})(2\pi k) = 0, \quad \textit{for all } k \in \mathbb{Z}^d \setminus \{0\},\ |\alpha| \leq r, \tag{3.3.8}$$

then we have the JACKSON *type inequality*

$$\inf_{g \in V_j} \|f - g\|_{L_p(\mathbb{R}^d)} \lesssim \|M^{-j}\|_\infty^{r+1} |f|_{W^{r+1}(L_p(\mathbb{R}^d))}. \tag{3.3.9}$$

Proof. Because ϕ is $(\mathbf{a}, M)$-refinable and ℓ_2-stable we get by (3.1.2) and (3.1.6) that

$$\hat{\phi}(0) \neq 0.$$

Application of Corollary 3.2.5 ensures that $\Pi_r \subset \mathcal{S}(\phi)$. Finally, we have all assumptions in hand to utilize Lemma 3.3.2 to continue. That is

$$|f(x) - (T_M f)(x)| \lesssim \|M\|_\infty^{r+1-\frac{d}{p}} |f|_{W^{r+1}(L_p(\|M\|_\infty c([-1,1]^d)+x))}.$$

We take the norm and raise to the power p

$$\begin{aligned}
\|f-T_Mf\|^p_{L_p(\mathbb{R}^d)} &= \int_{\mathbb{R}^d} |f(x)-T_Mf(x)|^p \mathrm{d}x \\
&\lesssim \|M\|_\infty^{(r-\frac{d}{p})p} \int_{\mathbb{R}^d} |f|^p_{W^{r+1}(L_p(\|M\|_\infty c([-1,1]^d)+x))} \mathrm{d}x \\
&\lesssim \|M\|_\infty^{(r+1-\frac{d}{p})p} \sum_{l\in\mathbb{Z}^d} \int_{\{\|M\|_\infty c([-1,1]^d)+\|M\|_\infty cl\}} \\
&\quad \left(\sum_{|\alpha|=r+1} \|\partial^\alpha f\|_{L_p(\|M\|_\infty c([-1,1]^d)+x)} \right)^p \mathrm{d}x \\
&\lesssim \|M\|_\infty^{(r+1-\frac{d}{p})p} \sum_{|\alpha|=r+1} \sum_{l\in\mathbb{Z}^d} \int_{\{\|M\|_\infty c([-1,1]^d)+\|M\|_\infty cl\}} \\
&\quad \|\partial^\alpha f\|^p_{L_p(\|M\|_\infty c([-1,1]^d)+x)} \mathrm{d}x \\
&\lesssim \|M\|_\infty^{(r+1)p} \sum_{|\alpha|=r+1} \sum_{l\in\mathbb{Z}^d} \|\partial^\alpha f\|^p_{L_p(\|M\|_\infty c([-1,1]^d)+\|M\|_\infty cl)} \\
&\lesssim \|M\|_\infty^{(r+1)p} \sum_{|\alpha|=r+1} \|\partial^\alpha f\|^p_{L_p(\mathbb{R}^d)}.
\end{aligned}$$

We get

$$\begin{aligned}
\|f-T_Mf\|_{L_p(\mathbb{R}^d)} &\lesssim \|M\|_\infty^{r+1} \left(\sum_{|\alpha|=r+1} \|\partial^\alpha f\|^p_{L_p(\mathbb{R}^d)} \right)^{\frac{1}{p}} \\
&\lesssim \|M\|_\infty^{r+1} |f|_{W^{r+1}(L_p(\mathbb{R}^d))}. \qquad (3.3.10)
\end{aligned}$$

Using a power of the matrix M in the definition of T_M in (3.3.3) we arrive at

$$(T_{M^{-j}}f)(x) = \sum_{k\in\mathbb{Z}^d} \tilde{L}(f(M^{-j}(\cdot+k)))\phi(M^jx-k).$$

Since we have

$$\inf_{g\in V_j} \|f-g\|_{L_p(\mathbb{R}^d)} \le \|f-T_{M^{-j}}f\|_{L_p(\mathbb{R}^d)}, \qquad (3.3.11)$$

we can apply inequality (3.3.10) to conclude

$$\inf_{g\in V_j} \|f-g\|_{L_p(\mathbb{R}^d)} \lesssim \|M^{-j}\|_\infty^{r+1} |f|_{W^{r+1}(L_p(\mathbb{R}^d))}, \qquad (3.3.12)$$

which shows the proposed JACKSON inequality. □

Chapter 4

Smoothness Properties of Shift-Invariant Spaces

Smoothness properties of shift-invariant spaces are strongly connected with so-called BERNSTEIN type estimates. Very often these types of estimate are also referred to as inverse estimates since they bound a stronger norm by a weaker one. In this chapter we investigate which smoothness order is achieved for functions in shift-invariant spaces. In particular we show several versions of BERNSTEIN type estimates which are basic ingredients for the general framework of approximation theory with wavelet decompositions as we will see later on in Chapter 5 when we take into account norm equivalences.

4.1 Bernstein Type Estimates

In this section we finally turn to smoothness properties of shift-invariant spaces. In particular, we derive sharp BERNSTEIN estimates that bound the smoothness norm of functions belonging to the spaces of a multiresolution analysis.

We start with a lemma that is similar to the formula of FAÁ DI BRUNO. The formula of FAÁ DI BRUNO gives an explicit equation for the $|\beta|$th derivative of the composition of two functions. See Roman (1980) for instance.

Lemma 4.1.1. *Let $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ be a function in $W^k(L_p(\mathbb{R}^d))$, $k \in \mathbb{N}$ and $A = (a_{ij})$ be a matrix in $\mathbb{Z}^{d\times d}$ that defines the mapping $x \longmapsto Ax$. Then we have for all $|\alpha| \le k$*

$$\frac{\partial^\alpha (f \circ A)}{\partial x^\alpha}(x) \le \|A\|_\infty^k \sum_{|\beta|=k} \frac{k!}{\beta!} \partial^\beta f(Ax). \tag{4.1.1}$$

Proof. This quite rough estimate can be obtained for instance by application of the formula of Faá di Bruno. Since we need to involve also the terms of the matrix norm $\|A\|_\infty$ we give here an alternative proof.

According to the chain rule we have for the Jacobian matrix $J_{f\circ A}(x) = J_f(Ax)$. Thus,

$$\frac{\partial (f \circ A)}{\partial x_j}(x) = \sum_{i=1}^d \frac{\partial f}{\partial y_i}(Ax) a_{ij}.$$

Iterated differentiation yields to

$$\begin{aligned}
\partial^\alpha f \circ A(x) &= \frac{\partial^\alpha f \circ A}{\partial x_{j_k} \ldots \partial x_{j_1}}(x) \\
&= \sum_{i_k,\ldots,i_1=1}^d \frac{\partial^\alpha f}{\partial y_{i_k} \ldots \partial y_{i_1}}(Ax) \cdot a_{i_1 j_1} \cdot \ldots \cdot a_{i_k j_k}.
\end{aligned}$$

We proceed with estimating the entries of the matrix A and get

$$\begin{aligned}
\partial^\alpha f \circ A(x) &\le \prod_{l=1}^k \max_{1\le i\le d} a_{ij_l} \sum_{i_k,\ldots,i_1=1}^d \frac{\partial^\alpha f}{\partial y_{i_k} \ldots \partial y_{i_1}}(Ax) \\
&= \prod_{l=1}^k \max_{1\le i\le d} a_{ij_l} \sum_{|\beta|=k} \frac{k!}{\beta_1! \cdot \ldots \cdot \beta_k!} \partial^\beta f(Ax) \\
&= \prod_{l=1}^d \left(\max_{1\le i\le d} a_{il} \right)^{\alpha_l} \sum_{|\beta|=k} \frac{k!}{\beta!} \partial^\beta f(Ax) \\
&\le \|A\|_\infty^k \sum_{|\beta|=k} \frac{k!}{\beta!} \partial^\beta f(Ax).
\end{aligned}$$

□

Our first version of a Bernstein type estimate relates Sobolev and L_p norms.

Theorem 4.1.2 (Bernstein inequality). *Let $\phi \in \mathcal{L}_p(\mathbb{R}^d) \cap W^k(L_p(\mathbb{R}^d))$ be a ℓ_p-stable function with $\partial^\alpha \phi \in \mathcal{L}_p(\mathbb{R}^d)$ for all $|\alpha| \le k$, $k \in \mathbb{N}$. Further, ϕ should generate a multiresolution analysis $V_0 = \operatorname{span}\{\phi(\cdot - k) :\ k \in \mathbb{Z}^d\}$, that is*

$$f(\cdot) \in V_j \iff f(\cdot) = g(M^j \cdot)\ , \quad g \in V_0.$$

Then, for all $f \in V_j$, $|\alpha| \le k$

$$\|\partial^\alpha f\|_{L_p(\mathbb{R}^d)} \lesssim \|M^j\|_\infty^k \|f\|_{L_p(\mathbb{R}^d)}. \tag{4.1.2}$$

Proof. First, we proof the case $j = 0$. We derive with the ℓ_p stability of ϕ and the Young type estimate (2.3.10)

$$\begin{aligned}
\|\partial^\alpha f\|_{L_p(\mathbb{R}^d)} &= \|\sum_{l \in \mathbb{Z}^d} b_l \partial^\alpha \phi(\cdot - l)\| \\
&= \|\partial^\alpha \phi *' b\|_{L_p(\mathbb{R}^d)} \\
&\lesssim |\partial^\alpha \phi|_p \|b\|_{\ell_p} \\
&\lesssim \|b\|_{\ell_p} \\
&\lesssim \|\phi *' b\|_{L_p(\mathbb{R}^d)} \\
&= \|f\|_{L_p(\mathbb{R}^d)}.
\end{aligned}$$

For $j \neq 0$ we apply (4.1.1) and observe that

$$\begin{aligned}
\|\partial^\alpha f\|_{L_p(\mathbb{R}^d)}^p &= \int_{\mathbb{R}^d} |\partial^\alpha g(M^j x)|^p \mathrm{d}x \\
&\le \int_{\mathbb{R}^d} \left| \|M^j\|_\infty^k \sum_{|\beta|=k} \frac{k!}{\beta!} \partial^\beta g(M^j x) \right|^p \mathrm{d}x \\
&= \|M^j\|_\infty^{kp} \int_{\mathbb{R}^d} |\det M|^{-j/2} \left| \sum_{|\beta|=k} \frac{k!}{\beta!} \partial^\beta g(x) \right|^p \mathrm{d}x \\
&\lesssim \|M^j\|_\infty^{kp} |\det M|^{-j/2} \int_{\mathbb{R}^d} |g(x)|^p \mathrm{d}x \\
&= \|M^j\|_\infty^{kp} \int_{\mathbb{R}^d} |g(M^j x)|^p \mathrm{d}x \\
&= \|M^j\|_\infty^{kp} \|f\|_{L_p(\mathbb{R}^d)}^p.
\end{aligned}$$

And we obtain

$$\|\partial^\alpha f\|_{L_p(\mathbb{R}^d)}^p \lesssim \|M^j\|_\infty^k \|f\|_{L_p(\mathbb{R}^d)}.$$

$\square$

To build up the general theory of norm equivalences between discrete smoothness norms and weighted sequences of wavelet decompositions we have to establish a more sophisticated version. This inequality is the main result of this chapter and involves the modulus of smoothness.

Theorem 4.1.3 (General Bernstein inequality). *Let $1 \leq p \leq \infty$. Assume that $\phi \in L_p(\mathbb{R}^d)$ and $\tilde{\phi} \in L_{p'}(\mathbb{R}^d)$, where $1 = \frac{1}{p} + \frac{1}{p'}$, are $(\mathbf{a}, M)$-refinable functions. Further, let $\phi, \tilde{\phi} \in W^{k+1}(\mathbb{R}^d)$. Then there exists a real number $\delta > 0$ such that for each $g \in V_m$*

$$\omega_{k+1}(g; t)_p \lesssim [\min\{1, t\rho^m\}]^{k+\delta} \|g\|_{L_p(\mathbb{R}^d)}. \tag{4.1.3}$$

Proof. Let $t < \rho^{-m}$, since otherwise this estimate is trivial. If we use the monotonicity of $\omega_{k+1}(g, t)_p$ in t it is sufficient to show the assertion for $t = \rho^{-l}$ with $l > m$. We decompose g in a local scaling function basis and consider the modulus of smoothness on the cubes $I_{m,k} = \|M^{-m}\|_\infty([0, 1]^d + k)$ as in the proof of the stability (Theorem 2.4.1),

$$\omega_{k+1}(g; t)_p \lesssim \omega_{k+1}(\phi_{m,0}; t)_p \|(c_k)_{k \in \mathbb{Z}^d}\|_{\ell_p}.$$

By the dilation property (2.5.30) of $\omega_{k+1}(g; t)_p$ it follows

$$\begin{aligned} \omega_{k+1}(\phi_{m,0}; \rho^{-l})_p &\lesssim |\det M|^{m(\frac{1}{2} - \frac{1}{p})} \omega_{k+1}(\phi; \rho^{m-l})_p \\ &\lesssim |\det M|^{m(\frac{1}{2} - \frac{1}{p})} \rho^{(m-l)(k+1)} |\phi|_{W^{k+1}(L_p(\mathbb{R}^d))} \\ &\lesssim |\det M|^{m(\frac{1}{2} - \frac{1}{p})}, \end{aligned}$$

where we have used (2.5.29), (2.5.31) and $l > m$. To conclude the proof we employ the lower inequality of the stability property (2.4.2). $\square$

Chapter 5

Besov Spaces and Wavelet Expansions

In this chapter we show the connection between approximation and smoothness properties. In particular we are interested which smoothness order of a function f ensures a given approximation rate by the shift-invariant spaces V_j and vice versa.

To be precise with the term approximation rate we introduce the spaces $\mathcal{A}_q^s(X)$. For a given function f we employ the *error of best approximation* of f by elements from the spaces approximation space X

$$\operatorname{dist}_Y(f, X) := \inf_{g \in X} \|f - g\|_Y. \tag{5.0.1}$$

Definition 5.0.1 (Approximation Spaces). *Let X be a* Banach *space and $(V_j)_{j\geq 0}$ a nested sequence of subspaces of X such that their union $\bigcup_{j\geq 0} V_j$ is dense in X. For real numbers $s > 0, \rho > 1$ and $1 \leq q \leq \infty$ we define the approximation space*

$$\mathcal{A}_q^s(X) := \{f \in X : \left(\sum_{j=0}^{\infty} (\rho^{sj} \operatorname{dist}_X(f, V_j))^q\right)^{1/q} < \infty\}. \tag{5.0.2}$$

The idea is that $\mathcal{A}_q^s(X)$ describes those functions that nearly satisfy $\operatorname{dist}_X(f, V_j) \leq \mathcal{O}(\rho^{-sj})$. It can be shown that $\mathcal{A}_q^s(X)$ constitutes a proper subspace of X. Also, it is a Banach space with the norm

$$\|f\|_{\mathcal{A}_q^s(X)} := \|f\|_X + |f|_{\mathcal{A}_q^s(X)}, \tag{5.0.3}$$

where the seminorm is given by

$$|f|_{\mathcal{A}_q^s(X)} := \left(\sum_{j=0}^{\infty} (\rho^{sj} \operatorname{dist}_X(f, V_j))^q \right)^{1/q} . \tag{5.0.4}$$

In the next section we show that

$$\|f\|_{\mathcal{A}_q^s(L_p(\mathbb{R}^d))} \sim \|f\|_{B_q^s(L_p(\mathbb{R}^d))}.$$

This identifies that Besov spaces are approximation spaces for non-separable wavelet expansions

$$\mathcal{A}_q^s(L_p(\mathbb{R}^d)) = B_q^s(L_p(\mathbb{R}^d)).$$

5.1 Characterization of Smoothness Spaces

Wavelet coefficients provide simple characterizations of most function spaces, in particular smoothness spaces. The norm in the given function space is equivalent to a sequence norm applied to the wavelet coefficients.

The first works with this topic date back to Meyer (1992) and Daubechies (1992). Valuable results with a whole theory behind was established e.g. by DeVore (1998) who first worked out this theory for dyadic spline approximation in DeVore and Popov (1988). A good introduction can be found in DeVore and Lucier (1992) or DeVore, Jawerth, and Popov (1992). For a recent overview with comprehensive explanations the reader is referred to Cohen (2003).

The well known results about approximation of functions by wavelet expansions and their related characterization of smoothness spaces such as Besov spaces are so far restricted to the dyadic case. In this chapter we provide such a characterization for L_p-spaces and Besov spaces for wavelet expansions with more general scaling matrices. We adhere to the ideas of Kunoth (1994) and Cohen (2003). In particular, the generalization consists of the choice of the scaling matrix M which is taken into account by means of its spectral norm $\rho = \rho(M) = \max_{i=1,\dots,d} |\lambda_i|$ with λ_i denoting the eigenvalues of the matrix M.

In this section we always consider functions defined on $\mathbb{R}^d$. The sequence of nested subspaces $\{V_j\}_{j\geq 0}$ shall establish a multiresolution analysis in terms of Definition 2.1.2. In what follows we assume that $j = 0$ is the coarsest level of the multiresolution analysis. The results for other values follow by rescaling. We emphasize that the multiresolution analysis is always associated with a $(\mathbf{h}, M)$-refinable function ϕ with an isotropic scaling matrix M and ρ denoting the modulus of its eigenvalues. The notations for the projections P_j, Q_j and the spaces V_j and W_j follow the line of the definitions of Section 2.2.

Further on, we assume that the wavelet basis $\Psi_j := \{\psi_{e,j,k},\ e \in E, k \in \mathbb{Z}^d\}$ constitutes a stable basis of W_j in the sense of (2.4.7), that is

$$\left\| \sum_{e\in E} \sum_{k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k} \right\|_{L_p(\mathbb{R}^d)} \sim |\det M|^{j(\frac{1}{2}-\frac{1}{p})} \|d^{(j)}\|_{\ell_p}. \tag{5.1.1}$$

with constants independent of the level $j \in \mathbb{Z}$. The sequence $d^{(j)}$ consists of all wavelet coefficients on level j

$$d^{(j)} := \{d_\lambda :\ \lambda \in \nabla_j\}.$$

Now we introduce a quantity that relates the approximation properties of the linear projectors P_j to the modulus of smoothness.

Definition 5.1.1 **($\nu^{(p)}_{m,j}$).** *Let $1 \le p \le \infty$ and $\{P_j\}_{j\in\mathbb{Z}}$ be a family of projection of a multiresolution analysis in terms of Definition 2.1.2 associated with a $(\mathbf{h}, M)$-refinable function ϕ with an isotropic scaling matrix M and ρ denoting the modulus of its eigenvalues. We define*

$$\nu^{(p)}_{m,j} := \sup_{f\in V_m} \frac{\|P_j f - f\|_{L_p(\Omega)}}{\omega_{r+1}(f, \rho^{-j})_p} \tag{5.1.2}$$

$$\nu^{(p)}_{m} := \max_{j=0,\ldots,m} \nu^{(p)}_{m,j} \tag{5.1.3}$$

Following the idea of approximation spaces we define a norm that controls the size of the wavelet coefficients.

Definition 5.1.2 (Multilevel Norm). *Let $\alpha \in \mathbb{R}^+$, $1 \le p, q \le \infty$, $m \in \mathbb{N}$ and $\{P_j\}_{j\in\mathbb{Z}}$ be a family of projection of a multiresolution analysis in terms of Definition 2.1.2 associated with a $(\mathbf{h}, M)$-refinable function ϕ with an isotropic scaling matrix M and ρ denoting the modulus of its eigenvalues. With $P_{-1} := 0$ we define for a function $f \in L_p(\mathbb{R}^d)$ its multilevel norm by*

$$\|f\|_{(\alpha,m,p,q)} := \left(\sum_{j=0}^{m} \rho^{j\alpha q} \|(P_j - P_{j-1})f\|^q_{L_p(\mathbb{R}^d)} \right)^{\frac{1}{q}}. \tag{5.1.4}$$

Next, following the line of DeVore and Popov (1988), we introduce a weighted sequence norm that we need in one of the theorems below to bound the BESOV norm.

Definition 5.1.3 ($\|a\|_{\ell^r_{q,\theta}}$ Norm). *For any sequence $a = (a_j)_{j\in\mathbb{N}_0}$, $1 < q \leq \infty$ and any real numbers $r > 0$, $\theta > 1$ we define*

$$\|a\|_{\ell^r_{q,\theta}} := \left(\sum_{j=0}^{\infty} \theta^{rjq} |a_j|^q \right)^{\frac{1}{q}}. \tag{5.1.5}$$

Similar to the idea in Dahmen and Micchelli (1983) we employ a discrete version of a HARDY type inequality. In contrast to the inequality there we provide a generalized variant of it. In particular, we keep more free parameters. This generalization is necessary since we deal with weighted BESOV norms containing more parameters. Especially the dilation 2 from the dyadic case is replaced by the weight ρ.

Lemma 5.1.1 (Discrete Hardy inequality). *Let $a = (a_j)_{j\in\mathbb{N}_0}$ be a given sequence. If for another sequence $b = (b_j)_{j\in\mathbb{N}_0}$ **one** of the following conditions*

$$(i) \quad |b_n| \lesssim B^{-n\delta} \left(\sum_{j=0}^{n} B^{j\delta} |a_j| \right), \quad B \geq 1, \ \delta > r > 0 \tag{5.1.6}$$

or

$$(ii) \quad |b_n| \lesssim \sum_{j=n}^{\infty} |a_j| \tag{5.1.7}$$

holds, then

$$\|b\|_{\ell^r_{q,\theta}} \lesssim \|a\|_{\ell^r_{q,\theta}}. \tag{5.1.8}$$

Proof. Let inequality (5.1.6) hold. Then, for some $\epsilon > 0$ such that $\delta - \epsilon > r$ we apply HÖLDER's inequality for $\frac{1}{q} + \frac{1}{\eta} = 1$, $1 < q, \eta < \infty$ and each $n \in \mathbb{N}_0$.

$$
\begin{aligned}
|b_n|^q &\lesssim B^{-n\delta q} \left(\sum_{j=0}^{n} B^{j\delta} |a_j| \right)^q \\
&= B^{-n\delta q} \left(\sum_{j=0}^{n} B^{j\epsilon} B^{j(\delta-\epsilon)} |a_j| \right)^q \\
&\lesssim B^{-n\delta q} \left(\sum_{j=0}^{n} B^{j\epsilon\eta} \right)^{\frac{q}{\eta}} \left(\sum_{j=0}^{n} B^{j(\delta-\epsilon)q} |a_j|^q \right) \\
&= B^{-n\delta q} \left(\frac{B^{\epsilon\eta(n+1)} - 1}{B^{\epsilon\eta} - 1} \right)^{\frac{q}{\eta}} \left(\sum_{j=0}^{n} B^{j(\delta-\epsilon)q} |a_j|^q \right) \\
&= B^{n(\epsilon-\delta)q} \left(\frac{B^{\epsilon\eta} - B^{-n\epsilon\eta}}{B^{\epsilon\eta} - 1} \right)^{\frac{q}{\eta}} \left(\sum_{j=0}^{n} B^{j(\delta-\epsilon)q} |a_j|^q \right) \\
&\lesssim B^{n(\epsilon-\delta)q} \left(\sum_{j=0}^{n} B^{j(\delta-\epsilon)q} |a_j|^q \right),
\end{aligned}
$$

where we used the sum formula for the geometric sums and the fact $B^{-n\epsilon\eta} < 1$. The constant depends on ϵ, η, q, B. Now we insert this inequality in $\sum_{n=0}^{\infty} \theta^{rnq} |b_n|^q$ and rearrange the sum. This yields

$$
\begin{aligned}
\sum_{n=0}^{\infty} \theta^{rnq} |b_n|^q &\lesssim \sum_{n=0}^{\infty} \theta^{rnq} B^{n(\epsilon-\delta)q} \left(\sum_{j=0}^{n} B^{j(\delta-\epsilon)q} |a_j|^q \right) \\
&= \sum_{n=0}^{\infty} B^{n(\delta-\epsilon)q} |a_n|^q \sum_{j=n}^{\infty} \theta^{rjq} B^{j(\epsilon-\delta)q} \\
&= \sum_{n=0}^{\infty} B^{n(\delta-\epsilon)q} |a_n|^q \frac{\theta^{rnq} B^{n(\epsilon-\delta)q}}{1 - \theta^{rq} B^{(\epsilon-\delta)q}} \\
&= \sum_{n=0}^{\infty} \theta^{rnq} |a_n|^q \frac{1}{1 - \theta^{rq} B^{(\epsilon-\delta)q}} \\
&\lesssim \sum_{n=0}^{\infty} \theta^{rnq} |a_n|^q,
\end{aligned}
$$

since $\epsilon - \delta + r < 0$. We used again the sum formula for geometric series. Thus, for the first case (5.1.6) we showed

$$\|b\|_{\ell^r_{q,\theta}} \lesssim \|a\|_{\ell^r_{q,\theta}}.$$

Now, assume (5.1.7) is valid. Let $\epsilon > 0$ and $1 < q, \eta < \infty$ such that $\frac{1}{q} + \frac{1}{\eta} = 1$. Then, we obtain for each $n \in \mathbb{N}_0$

$$\begin{aligned} |b_n|^q &\lesssim \left(\sum_{j=n}^{\infty} |a_j| \right)^q \\ &= \left(\sum_{j=n}^{\infty} \theta^{j\epsilon} \theta^{-j\epsilon} |a_j| \right)^q \\ &\lesssim \left(\sum_{j=n}^{\infty} \theta^{-j\epsilon\eta} \right)^{\frac{q}{\eta}} \left(\sum_{j=n}^{\infty} \theta^{j\epsilon q} |a_j|^q \right) \\ &= \left(\frac{\theta^{-n\epsilon\eta}}{1 - \theta^{-n\epsilon\eta}} \right)^{\frac{q}{\eta}} \left(\sum_{j=n}^{\infty} \theta^{j\epsilon q} |a_j|^q \right) \\ &\lesssim \theta^{-n\epsilon q} \left(\sum_{j=n}^{\infty} \theta^{j\epsilon q} |a_j|^q \right). \end{aligned}$$

Finally we insert this inequality in $\sum_{n=0}^{\infty} \theta^{rnq} |b_n|^q$ and get

$$\begin{aligned} \sum_{n=0}^{\infty} \theta^{rnq} |b_n|^q &\lesssim \sum_{n=0}^{\infty} \theta^{rnq} \theta^{-n\epsilon q} \left(\sum_{j=n}^{\infty} \theta^{j\epsilon q} |a_j|^q \right) \\ &= \sum_{n=0}^{\infty} \theta^{n\epsilon q} |a_n|^q \left(\sum_{j=0}^{n} \theta^{jq(r-\epsilon)} \right) \\ &= \sum_{n=0}^{\infty} \theta^{n\epsilon q} |a_n|^q \frac{\theta^{q(r-\epsilon)(n+1)} - 1}{\theta^{q(r-\epsilon)} - 1} \\ &= \sum_{n=0}^{\infty} \theta^{r\epsilon q} |a_n|^q \frac{\theta^{(r-\epsilon)q} - \theta^{-(r-\epsilon)nq}}{\theta^{(r-\epsilon)q} - 1} \\ &\lesssim \sum_{n=0}^{\infty} \theta^{rnq} |a_n|^q, \end{aligned}$$

where the last inequality is valid because $\theta^{-(r-\epsilon)nq} < 1$. $\square$

Now, we are prepared to state a result which assures the desired norm equivalence between discrete BESOV space norms and weighted sequences of wavelet coefficients. The following theorem demonstrates the dependence of these norm equivalences on BERNSTEIN or *inverse inequalities.*

Theorem 5.1.2. *Suppose that for some real number $\gamma > r$ and for all $m \in \mathbb{N}_0$ the* BERNSTEIN *inequality*

$$\omega_{r+1}(g;t)_p \lesssim (\min\{1, t\rho^m\})^\gamma \|g\|_{L_p(\mathbb{R}^d)} \quad \textit{for } g \in V_m \tag{5.1.9}$$

is valid. Then, for each $0 < \alpha < \min\{\gamma, r+1\}$

$$\frac{\|g\|_{(\alpha,m,p,q)}}{\nu_m^{(p)}} \lesssim \|g\|_{B_q^\alpha(L_p(\mathbb{R}^d))} \lesssim \|g\|_{(\alpha,m,p,q)} \quad \textit{for any } g \in V_m. \tag{5.1.10}$$

Proof. We show

(i) $\frac{\|g\|_{(\alpha,m,p,q)}}{\nu_m^{(p)}} \lesssim \|g\|_{B_q^\alpha(L_p(\mathbb{R}^d))}$

(ii) $\|g\|_{B_q^\alpha(L_p(\mathbb{R}^d))} \lesssim \|g\|_{(\alpha,m,p,q)}$

Let us first show item (i). We use a telescopic sum to obtain

$$\begin{aligned} \|g\|_{(\alpha,m,p,q)} &= \left(\sum_{j=0}^{m} \rho^{j\alpha q} \|(P_j - P_{j-1})g\|_{L_p(\mathbb{R}^d)}^q \right)^{\frac{1}{q}} \\ &\lesssim \|g\|_{L_p(\mathbb{R}^d)} + \left(\sum_{j=0}^{m} \rho^{j\alpha q} \|P_j g - g\|_{L_p(\mathbb{R}^d)}^q \right)^{\frac{1}{q}}. \end{aligned}$$

By definition of $\nu_m^{(p)}$ and $\nu_{m,j}^{(p)}$ in (5.1.2) and (5.1.3) it follows that

$$\|P_j g - g\|_{L_p(\mathbb{R}^d)} \lesssim \nu_m^{(p)} \omega_{r+1}(g, \rho^{-j})_p \quad \text{for } 0 \le j \le m.$$

Thus, with the definition of the BESOV seminorm (2.5.35) and the equivalence of all discrete BESOV seminorms, we get

$$\begin{aligned} \|g\|_{(\alpha,m,p,q)} &\lesssim \|g\|_{L_p(\mathbb{R}^d)} + \nu_m^{(p)} \left(\sum_{j=0}^{m} \rho^{j\alpha q} \omega_{r+1}(g, \rho^{-j})_p^q \right)^{\frac{1}{q}} \\ &\lesssim \nu_m^{(p)} \left(\|g\|_{L_p(\mathbb{R}^d)} + |g|_{B_q^\alpha(L_p(\mathbb{R}^d))} \right) \lesssim \nu_m^{(p)} \|g\|_{B_q^\alpha(L_p(\mathbb{R}^d))}. \end{aligned}$$

Now we consider item (ii). Assume that $g = \sum_{j=0}^{\infty} g_j$ for $g_j \in V_j$, $j \geq 0$. Then the subadditivity of the modulus of smoothness implies

$$\begin{aligned} \omega_{r+1}(g,\rho^{-n})_p &\lesssim \sum_{j=0}^{\infty} \omega_{r+1}(g_j,\rho^{-n})_p \\ &\lesssim \sum_{j=0}^{n} \rho^{(j-n)\gamma} \|g_j\|_{L_p(\mathbb{R}^d)} + \sum_{j=n+1}^{\infty} \|g_j\|_{L_p(\mathbb{R}^d)}, \end{aligned}$$

where the last inequality follows from the assumed Bernstein inequality (5.1.9).

If we set $l = r+1$ as the order of the modulus of smoothness in (2.5.35) it becomes

$$\begin{aligned} |g|_{B_q^\alpha(L_p(\mathbb{R}^d))} &\sim \left(\sum_{n=0}^{\infty} \rho^{n\alpha q} \omega_{r+1}(g,\rho^{-n})_p \right)^{\frac{1}{q}} \\ &\lesssim \left(\sum_{n=0}^{\infty} \rho^{n\alpha q} \left[\sum_{j=0}^{n} (\rho^{(j-n)\gamma} \|g_j\|_{L_p(\mathbb{R}^d)} + \sum_{j=n+1}^{\infty} \|g_j\|_{L_p(\mathbb{R}^d)} \right]^q \right)^{\frac{1}{q}} \\ &\lesssim \left(\sum_{n=0}^{\infty} \rho^{nq(\alpha-\gamma)} \left[\sum_{j=0}^{n} \rho^{j\gamma} \|g_j\|_{L_p(\mathbb{R}^d)} \right]^q + \right. \\ &\qquad \left. \rho^{nq\alpha} \left[\sum_{j=n+1}^{\infty} \|g_j\|_{L_p(\mathbb{R}^d)} \right]^q \right)^{\frac{1}{q}} \\ &\lesssim \left(\sum_{n=0}^{\infty} \rho^{n\alpha q} |b_{1,n}|^q \right)^{\frac{1}{q}} + \left(\sum_{n=0}^{\infty} \rho^{n\alpha q} |b_{2,n}|^q \right)^{\frac{1}{q}}, \end{aligned}$$

with

$$b_{1,n} := \rho^{-n\gamma} \sum_{j=0}^{n} \rho^{j\gamma} \|g_j\|_{L_p(\mathbb{R}^d)}$$

and

$$b_{2,n} := \sum_{j=n+1}^{\infty} \|g_j\|_{L_p(\mathbb{R}^d)}.$$

Thus we can simplify to

$$|g|_{B_q^\alpha(L_p(\mathbb{R}^d))} \lesssim \|(b_{1,n})_{n\in\mathbb{N}}\|_{\ell_{q,\rho}^\alpha} + \|(b_{2,n})_{n\in\mathbb{N}}\|_{\ell_{q,\rho}^\alpha}. \tag{5.1.11}$$

Since $\gamma > \alpha$ and $\rho > 1$ we can apply the discrete HARDY inequality (5.1.8) and conclude

$$\|b_{1,\cdot}\|_{\ell^{\alpha}_{q,\rho}} + \|b_{2,\cdot}\|_{\ell^{\alpha}_{q,\rho}} \lesssim \left(\sum_{n=0}^{\infty} \rho^{n\alpha q}\|g_n\|^q_{L_p(\mathbb{R}^d)}\right)^{\frac{1}{q}}.$$

The particular choice $g_n = (P_n - P_{n-1})g$ brings the right quantity to $\|g\|_{(\alpha,\infty,p,q)}$ which is exactly what we wanted to show. □

To establish the lower inequality of the norm equivalence (5.1.10) we have to bound $\nu_m^{(p)}$. This is achieved with the general JACKSON estimate that implies that $\nu_m^{(p)} = \mathcal{O}(1)$, as $m \to \infty$.

Theorem 5.1.3. *Let $P = \{P_j\}_{j\in\mathbb{N}_0}$ a uniformly bounded sequence of projectors in $L_p(\mathbb{R}^d)$ of a multiresolution analysis in terms of Definition 2.1.2 associated with a $(\mathbf{h}, M)$-refinable function ϕ with an isotropic scaling matrix M. If the following* JACKSON *estimate*

$$E(f, V_m)_p := \inf_{g\in V_m} \|f - g\|_{L_p} \lesssim \|M^{-m}\|^l_\infty |f|_{W^l(L_p)} \tag{5.1.12}$$

is valid for $l > r$, (where $r+1$ is the order of the modulus of smoothness in (5.1.2)) then

$$\nu_m^{(p)} = \mathcal{O}(1), \qquad \text{as } m \to \infty. \tag{5.1.13}$$

Proof. For $f \in W^l(L_p(\mathbb{R}^d))$ and $g \in V_m$ one has $\|f-g\|_{L_p} \le \|f-h\|_{L_p} + \|g-h\|_{L_p}$. Thus, if $h \in W^l(L_p(\mathbb{R}^d))$ with (5.1.12) we obtain

$$\begin{aligned} E(f, V_m)_p &\lesssim \|f-h\|_{L_p} + E(h, V_m)_p \\ &\lesssim \|f-h\|_{L_p} + \|M^{-m}\|^l_\infty |h|_{W^l(L_p)}. \end{aligned}$$

We employ the so called PEETRE *K-functional.* Defining

$$K_l(f,t) := \inf_{h\in W^l(L_p(\mathbb{R}^d))} \|f-h\|_{L_p} + t|h|_{W^l(L_p)} \tag{5.1.14}$$

we recall from Johnen and Scherer (1977) that

$$K_l(f, t^l) \sim \omega_l(f,t)_p \quad \text{as} \quad \text{for } t > 0 \tag{5.1.15}$$

and conclude

$$E(f, V_m)_p \lesssim \omega_l(f, \|M^{-m}\|_\infty)_p. \tag{5.1.16}$$

Now we perform a LEBESGUE type estimate. For any $h \in V_m$ we derive

$$\begin{aligned}\|P_m f - f\|_{L_p} &\leq \|P_m f - P_m h\|_{L_p} + \|h - f\|_{L_p} \\ &\leq (\|P_m\|_p + 1)\|h - f\|_{L_p} \\ &\lesssim (\|P_m\|_p + 1)E(f, V_m)_p \\ &\lesssim (\|P_m\|_p + 1)\omega_l(f, \|M^{-m}\|_\infty)_p,\end{aligned}$$

where the last inequality follows from (5.1.16). Since $h \in V_m$ is arbitrary the inequality holds in particular for the infimum with respect to the L_p-norm, which finally yields

$$\nu_m^{(p)} = \mathcal{O}(1).$$

□

Remark 5.1.1. *The quantity* $\|f - P_j f\|_{L_p(\mathbb{R}^d)}$ *is equivalent to* $\mathrm{dist}_p(f, V_j)$ *for all* $1 \leq p \leq \infty$. *Since* $P_j f \in V_j$ *we obviously have*

$$\mathrm{dist}_p(f, V_j) \leq \|f - P_j f\|_{L_p(\mathbb{R}^d)}.$$

To find the converse inequality we obtain for any $g \in V_j$

$$\begin{aligned}\|f - P_j f\|_{L_p(\mathbb{R}^d)} &\leq \|f - g\|_{L_p(\mathbb{R}^d)} + \|P_j f - P_j g\|_{L_p(\mathbb{R}^d)} \\ &\leq (1 + \|P_j\|)\,\mathrm{dist}_p(f, V_j),\end{aligned}$$

which is finite as equation (5.1.13) shows.

5.2 Norm Equivalences and Linear Approximation

Now we are ready to state the main result about norm equivalences between non-separable wavelet expansions and discrete BESOV norms.

Theorem 5.2.1 (Norm Equivalences). *Let* ϕ *be a* $(\mathbf{h}, M)$*-refinable function associated with a multiresolution analysis in terms of Definition 2.1.2 with an isotropic scaling matrix* M *and* ρ *denoting the modulus of its eigenvalues and* $\mathbf{h} \in \ell_1(\mathbb{Z}^d)$. *Furthermore, let the assumptions of the stability theorem 2.4.1, the* JACKSON *inequality 3.3.3 and the* BERNSTEIN *inequality 4.1.3 hold.*

Then, for $0 < \alpha < k + \delta$, $\delta > 0$ small enough, the following conditions are equivalent:

$$(i) \qquad \left(\sum_{j=-1}^{\infty} (\rho^{j\alpha} \|f - P_j f\|_{L_p(\mathbb{R}^d)})^q \right)^{\frac{1}{q}} < \infty \tag{5.2.1}$$

$$(ii) \qquad \left(\sum_{j=-1}^{\infty} (\rho^{j\alpha} \|Q_j f\|_{L_p(\mathbb{R}^d)})^q \right)^{\frac{1}{q}} < \infty \tag{5.2.2}$$

$$(iii) \qquad \left(\sum_{k \in \mathbb{Z}^d} |\langle f, \tilde{\phi}(\cdot - k)\rangle|^p \right)^{\frac{1}{p}} + \left(\sum_{j=0}^{\infty} (\rho^{j\alpha} |\det M|^{j(\frac{1}{2}-\frac{1}{p})} \|d^{(j)}\|_{\ell_p})^q \right)^{\frac{1}{q}} < \infty \tag{5.2.3}$$

$$(iv) \qquad f \in \mathcal{A}_q^{\alpha}(L_p(\mathbb{R}^d)) \tag{5.2.4}$$

$$(v) \qquad \|f\|_{(\alpha,\infty,p,q)} < \infty \tag{5.2.5}$$

$$(vi) \qquad f \in B_q^{\alpha}(L_p(\mathbb{R}^d)). \tag{5.2.6}$$

Proof. The equivalence (5.2.1)$\Longleftrightarrow$(5.2.4) is clear by Remark 5.1.1.

Also, by definition of the projections $Q_j = P_{j+1} - P_j$ we see the equivalence (5.2.2)$\Longleftrightarrow$(5.2.5).

For the important fact (5.2.2)$\Longleftrightarrow$(5.2.3) it suffices to recall the stability properties (2.4.2) and (2.4.7) of the projectors P_j and Q_j which ensure $\|P_j\|_{L_p(\mathbb{R}^d)} < C_1$ and $\|Q_j\|_{L_p(\mathbb{R}^d)} < C_2$ for positive constants C_1, C_2 independently of j. Using the notation

$$d^{(j)} := \{d_\lambda : \ \lambda \in \nabla_j\}.$$

proves the equivalence.

To show (5.2.5)$\Longleftrightarrow$(5.2.6) we use Theorem 4.1.3 which implies that the Bernstein inequality (5.1.9) is valid for $r = k$ and $\gamma = k + \delta$. This shows that inequality (5.1.10) holds. Theorem 3.3.3 ensures that for isotropic scaling matrices the Jackson type estimate (5.1.12) is valid with $l = k + 1$ which bounds $\nu_m^{(p)}(f)$ in (5.1.13).

Since $Q_j(f - P_j f) = (P_{j+1} - P_j)f = Q_j f$ and $\|Q_j\|_{L_p(\mathbb{R}^d)} \le C$ for a constant C independently of j we conclude $\|Q_j f\|_{L_p(\mathbb{R}^d)} \le C\|f - P_j f\|_{L_p(\mathbb{R}^d)}$ and (5.2.1)$\Longrightarrow$(5.2.2).

For the final implication (5.2.2)$\Longrightarrow$(5.2.1) we first estimate

$$\sum_{j=0}^{\infty} \rho^{j\alpha q} \|f - P_j f\|^q_{L_p(\mathbb{R}^d)} \leq \sum_{j=0}^{\infty} \rho^{j\alpha q} (\sum_{m=j}^{\infty} \|(P_{m+1} - P_m)f\|_{L_p(\mathbb{R}^d)})^q. \quad (5.2.7)$$

Then we take $0 < a < \alpha$ and use HÖLDER's inequality with $\frac{1}{q} + \frac{1}{q'} = 1$ to obtain that the right hand side of (5.2.7) is bound by

$$\begin{aligned}
&\sum_{j=0}^{\infty} \rho^{j\alpha q} (\sum_{m=j}^{\infty} \rho^{-am} \rho^{am} \|(P_{m+1} - P_m)f\|_{L_p(\mathbb{R}^d)})^q \\
&\qquad \leq \sum_{j=0}^{\infty} \rho^{j\alpha q} (\sum_{m=j}^{\infty} \rho^{-amq'})^{q/q'} (\sum_{m=j}^{\infty} \rho^{amq} \|(P_{m+1} - P_m)f\|^q_{L_p(\mathbb{R}^d)}) \\
&\qquad \lesssim \sum_{j=0}^{\infty} \rho^{j\alpha q} \rho^{-j\alpha q} \sum_{m=j}^{\infty} \rho^{amq} \|(P_{m+1} - P_m)f\|^q_{L_p(\mathbb{R}^d)} \\
&\qquad = \sum_{m=0}^{\infty} \rho^{amq} \|(P_{m+1} - P_m)f\|^q_{L_p(\mathbb{R}^d)} \sum_{j \leq m} \rho^{jq(\alpha - a)}. \qquad (5.2.8)
\end{aligned}$$

Summation of geometric series shows that the right hand side of (5.2.8) is bound by

$$\sum_{m=0}^{\infty} \rho^{amq} \rho^{mq(\alpha - a)} \|(P_{m+1} - P_m)f\|^q_{L_p(\mathbb{R}^d)} = \sum_{m=0}^{\infty} \rho^{mq\alpha} \|(P_{m+1} - P_m)f\|^q_{L_p(\mathbb{R}^d)},$$

which concludes the proof. □

For a particular choice of p and q we get a nice characterization of fractional SOBOLEV spaces which is specified by the following corollary.

Corollary 5.2.2. *Under the assumptions of Theorem 5.2.1 and the additional requirements that $\tilde{\phi} \in \mathcal{L}_2$ and $\operatorname{supp} \tilde{\phi}$ is compact we get for any*

$$\begin{aligned}
f &= \sum_{k \in \mathbb{Z}^d} c_k \phi(\cdot - k) + \sum_{e \in E} \sum_{j \geq 0} \sum_{k \in \mathbb{Z}^d} d_{e,j,k} \psi_{e,j,k} \qquad (5.2.9) \\
&= \sum_{k \in \mathbb{Z}^d} \langle f, \tilde{\phi}(\cdot - k) \rangle \phi(\cdot - k) + \sum_{e \in E} \sum_{j \geq 0} \sum_{k \in \mathbb{Z}^d} \langle f, \tilde{\psi}_{e,j,k} \rangle \psi_{e,j,k}
\end{aligned}$$

the following equivalence:

$$f \in H^\alpha(\mathbb{R}^d) \Longleftrightarrow \|(c_k)_{k \in \mathbb{Z}^d}\|^2_{\ell_2} + \sum_{j \geq 0} (\rho^{j\alpha} \|d^{(j)}\|_{\ell_2})^2 < \infty, \qquad (5.2.10)$$

where

$$d^{(j)} := \{d_\lambda : \ \lambda \in \nabla_j\}.$$

Proof. Since $\operatorname{supp} \tilde{\phi}$ is compact we easily obtain by Theorem 2.4.1 that for any $f \in L_2(\mathbb{R}^d)$ the family $\{P_j\}_{j \geq 0}$ is uniformly bounded:

$$\|P_j f\|_{L_2(\mathbb{R}^d)} \lesssim \|P_j\|_{L_2(\mathbb{R}^d)} \|f\|_{L_2(\mathbb{R}^d)}.$$

We employ the norm equivalences of Theorem 5.2.1

$$\begin{aligned}
\|f\|^2_{H^\alpha(\mathbb{R}^d)} &= \|f\|_{B^\alpha_2(L_2(\mathbb{R}^d))} \sim \|f\|^2_{(\alpha,\infty,2,2)} \\
&\sim \sum_{j=0}^{\infty} (\rho^{j\alpha} \|P_j f - P_{j-1} f\|_{L_2(\mathbb{R}^d)})^2 \\
&\sim \|P_0 f\|^2_{L_2(\mathbb{R}^d)} + \sum_{j=1}^{\infty} (\rho^{j\alpha} \|P_j f - P_{j-1} f\|_{L_2(\mathbb{R}^d)})^2 \\
&\sim \|P_0 f\|^2_{L_2(\mathbb{R}^d)} + \sum_{j=0}^{\infty} (\rho^{j\alpha} \|d^{(j)}\|_{\ell_2})^2.
\end{aligned}$$

The stability of the projections P_j concludes the proof. □

Remark 5.2.1. *We have seen that approximation order α in $L_p(\mathbb{R}^d)$ can be characterized by smoothness of order α measured in the same norm. This type of approximation is associated with linear approximation since the approximation spaces V_j are linear spaces. A visualization is depicted in Figure 5.1.*

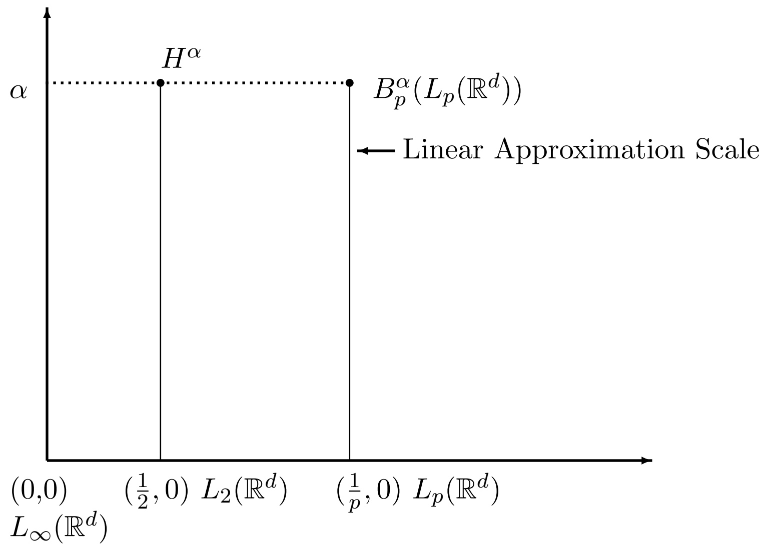


Figure 5.1: Graphical interpretation of the approximation scale for linear approximation.

Chapter 6

Negative Smoothness and Approximation

In the last chapter we have considered Besov spaces with smoothness index s strictly larger than 0. A typical setting for partial differential equations is an operator that maps a function from a smoothness space of order s into another of order $-s$. For the treatment of partial differential equations with negative smoothness of the solution, e.g. numerical wavelet Galerkin schemes, it is therefore essential to have also a characterization for Besov spaces with negative smoothness.

Subject matter of this chapter is the characterization of Besov spaces via weighted discrete norms of wavelet decompositions for smoothness order s in the range $s < 0$.

6.1 Besov Spaces with Negative Smoothness by Means of Duality

We define the Besov space $B_q^{-s}(L_p(\mathbb{R}^d))$ as the set of all functions f with finite norm

$$\|f\|_{B_q^{-s}(L_p(\mathbb{R}^d))} := \left(\sum_{j \geq 0} (\rho^{-sj} \omega_l(f, \rho^{-j})_p)^q \right)^{1/q}, \qquad j \in \mathbb{N},\ l > s. \quad (6.1.1)$$

In general one identifies Besov spaces $B_q^{-s}(L_p(\mathbb{R}^d))$ for $s > 0$ by duality as

$$B_{q'}^{-s}(L_{p'}(\mathbb{R}^d)) = (B_q^s(L_p(\mathbb{R}^d)))' \quad (6.1.2)$$

with $1 \leq p, q \leq \infty$ such that $\frac{1}{p} + \frac{1}{p'} = 1 = \frac{1}{q} + \frac{1}{q'}$. In this context we employ again the biorthogonal setting, that is, we characterize the dual BESOV space by means of the wavelet decomposition of the dual wavelet basis. This idea was designed in a general sense for shift-invariant spaces in Dahmen (1996) and Dahmen (1995). A particular proof for dyadic wavelet decompositions was done in Kunoth (1994, Sec. 3.1). We remember that we already used this concept where we characterized the primal BESOV space by the dual wavelet coefficients in Section 5.2, Theorem 5.2.1.

For a BANACH space $\mathcal{F}$ the dual is denoted by $\mathcal{F}'$. For any linear operator $P : \mathcal{F} \longmapsto \mathcal{F}$, its dual operator $\tilde{P}$ is defined as

$$\langle \tilde{P}\tilde{f}, f\rangle := \langle \tilde{f}, Pf\rangle, \tag{6.1.3}$$

with $f \in \mathcal{F}$, $\tilde{f} \in \mathcal{F}'$ where the duality $\langle f, g\rangle$ is meant in the sense of $(\mathcal{F}, \mathcal{F}')$.

The dual projectors to P_j and Q_j are

$$\tilde{P}_j := \sum_{k\in\mathbb{Z}^d} \langle f, \phi_{j,k}\rangle \tilde{\phi}_{j,k} \tag{6.1.4}$$

and

$$\tilde{Q}_j := \sum_{k\in\mathbb{Z}^d} \langle f, \psi_{j,k}\rangle \tilde{\psi}_{j,k}. \tag{6.1.5}$$

Remark 6.1.1. *In Dahmen (1994) the following properties are derived which we collect here since we need them shortly.*

- *Since P_j is a linear operator so is $\tilde{P}_j$. Thus, in comparison to (2.2.2) we also get*
$$\tilde{P}_j\tilde{P}_n = \tilde{P}_j, \qquad \textit{for } j \leq n. \tag{6.1.6}$$
- *The biorthogonality relation $W_j \perp \tilde{W}_k$ for $j \neq k$ is equivalent to*
$$\langle(\tilde{P}_j - \tilde{P}_{j-1})\tilde{f}, (P_k - P_{k-1})f\rangle = 0, \qquad \textit{for } j \neq k \tag{6.1.7}$$
for any $\tilde{f} \in \mathcal{F}'$, $f \in \mathcal{F}$.

Remark 6.1.2. *Note that for all $\rho \in \mathbb{R}$ all these norms are equivalent, cf. 2.5.34.*

In analogy to the definition of the sequence norm defined in (5.1.4) we introduce the dual version by

$$\|f\|^{\sim}_{(s,m,p',q')} := \left(\sum_{j=0}^{m} (\rho^{js} \|(\tilde{P}_j - \tilde{P}_{j-1})f\|_{L_{p'}(\mathbb{R}^d)})^{q'} \right)^{\frac{1}{q'}}, \tag{6.1.8}$$

with $1 < p' < \infty$ and $0 < q' < \infty$.

Now we show that this norm is equivalent to the sequence norm as defined in 6.1.8, but s replaced by $(-s)$. We follow the line of the proofs of Kunoth (1994, Sec. 3.1).

Theorem 6.1.1. *Let $(\phi, \tilde{\phi})$ be a biorthogonal pair of $(\mathbf{h}, M)$-refinable (or $(\tilde{\mathbf{h}}, M)$-refinable resp.) functions as in (2.1.19) and (2.1.24) associated with a multiresolution analysis in terms of Definition 2.1.2 with an isotropic scaling matrix M and ρ denoting the modulus of its eigenvalues and $\mathbf{h} \in \ell_1(\mathbb{Z}^d)$. Further let $(\psi, \tilde{\psi})$ be a wavelet basis generated by $(\phi, \tilde{\phi})$.*

Suppose that for $k \in \mathbb{N}$ and some real number $\delta > k$ and for the sequences $\{V_j\}_{j\geq 0}$ and $\{\tilde{V}_j\}_{j\geq 0}$ the Bernstein *inequality (4.1.3) is valid relative to $L_p(\mathbb{R}^d)$, resp. $L_{p'}(\mathbb{R}^d)$, $1 \leq p, p' \leq \infty$. Furthermore, let the sequences $\{P_j\}_{j\geq 0}$ and $\{\tilde{P}_j\}_{j\geq 0}$ be uniformly bounded in $L_p(\mathbb{R}^d)$ and $L_{p'}(\mathbb{R}^d)$.*
Then for any $0 < s < \min\{\delta, k+1\}$ and each $f \in V_m$ one has

$$\|g\|^{\sim}_{(-s,m,p',q')} \sim \|g\|_{(B^s_q(L_p(\mathbb{R}^d)))'}, \tag{6.1.9}$$

with $1 < p, q, p', q' < \infty$ such that $\frac{1}{p} + \frac{1}{p'} = 1$ and $\frac{1}{q} + \frac{1}{q'} = 1$.

Proof. First, we show the lower inequality

$$\|g\|_{(B^s_q(L_p(\mathbb{R}^d)))'} \lesssim \|g\|^{\sim}_{(-s,m,p',q')}.$$

Using the multilevel representation we write for any $g \in \tilde{V}_m$ and $g \in B^s_q(L_p(\mathbb{R}^d))$

$$\begin{aligned} \langle g, f\rangle &= \left\langle \sum_{j=0}^{m} (\tilde{P}_j - \tilde{P}_{j-1})g, \sum_{k=0}^{\infty} (P_k - P_{k-1})f \right\rangle \\ &= \sum_{j=0}^{m}\sum_{k=0}^{\infty} \langle (\tilde{P}_j - \tilde{P}_{j-1})g, (P_k - P_{k-1})f\rangle \\ &= \sum_{j=0}^{m} \langle (\tilde{P}_j - \tilde{P}_{j-1})g, (P_j - P_{j-1})f\rangle, \end{aligned}$$

where we have used the properties (6.1.7) and (6.1.6) . If we now apply first a discrete Hölder inequality and afterwards the continuous version we can bound $|\langle g, f\rangle|$ by

$$\begin{aligned}\sum_{j=0}^{m} &\rho^{-sj}\|(\tilde{P}_j-\tilde{P}_{j-1})g\|_{L_{p'}(\mathbb{R}^d)}\rho^{sj}\|(P_j-P_{j-1})f\|_{L_p(\mathbb{R}^d)}\\ &\le \left(\sum_{j=0}^{m}(\rho^{-sj}\|(\tilde{P}_j-\tilde{P}_{j-1})g\|_{L_{p'}(\mathbb{R}^d)})^{q'}\right)^{1/q'}\\ &\qquad\cdot\left(\sum_{j=0}^{m}(\rho^{sj}\|(P_j-P_{j-1})f\|_{L_p(\mathbb{R}^d)})^{\xi}\right)^{1/\xi}\\ &= \|g\|^{\sim}_{(-s,m,p',q')}\|f\|_{(s,m,p,\xi)},\end{aligned}$$

where $\frac{1}{p}+\frac{1}{p'}=\frac{1}{\xi}+\frac{1}{q'}=1$.

In the sense of duality the norm is defined as

$$\|f\|_{B^s_{q'}(L_{p'}(\mathbb{R}^d))} := \sup_{\substack{g\in B^s_q(L_p(\mathbb{R}^d))\\ \|g\|_{B^s_q(L_p(\mathbb{R}^d))}=1}} |\langle f, g\rangle| = \sup_{g\in B^s_q(L_p(\mathbb{R}^d))}\frac{|\langle g, f\rangle|}{\|g\|_{B^s_q(L_p(\mathbb{R}^d))}}. \tag{6.1.10}$$

Thus we arrive at

$$\|g\|_{(B^s_q(L_p(\mathbb{R}^d)))'} \le \sup_{f\in B^s_q(L_p(\mathbb{R}^d)), f\neq 0}\frac{\|g\|^{\sim}_{(-s,m,p',q')}\|f\|_{(s,m,p,\xi)}}{\|f\|_{B^s_q(L_p(\mathbb{R}^d))}}. \tag{6.1.11}$$

Now we employ the equivalence (5.1.10) to show that $\frac{\|f\|_{(s,m,p,\xi)}}{\|f\|_{B^s_q(L_p(\mathbb{R}^d))}}$ is bounded. Since we assumed that the sequences $\{P_j\}_{j\ge 0}$ and $\{\tilde{P}_j\}_{j\ge 0}$ are uniformly bounded in $L_p(\mathbb{R}^d)$ and $L_{p'(\mathbb{R}^d)}$ this implies that $\nu_m^{(p)}=\mathcal{O}(1)$. Therefore the terms involving f in the last inequality cancel down to a constant for $\xi=q$. Hence we have proved the upper inequality of (6.1.9).

To show the lower inequality of (6.1.9)

$$\|g\|^{\sim}_{(-s,m,p',q')} \lesssim \|g\|_{(B^s_q(L_p(\mathbb{R}^d)))'},$$

we fix an s such that $0<s<\min\{\delta, k+1\}$. For the operator

$$\tilde{\Theta}: g \longmapsto \sum_{l\ge 0}\rho^{-sl}(\tilde{P}_l-\tilde{P}_{l-1})g \tag{6.1.12}$$

we get with property (6.1.7)

$$(\tilde{P}_j - \tilde{P}_{j-1})\tilde{\Theta} g = \rho^{-sj}(\tilde{P}_j - \tilde{P}_{j-1})g. \tag{6.1.13}$$

Thus, we rewrite $\|g\|^{\sim}_{(-s,m,p',q')}$ and bound it as follows

$$\|g\|^{\sim}_{(-s,m,p',q')} = \left(\sum_{j=0}^{m}(\rho^{-js}\|(\tilde{P}_j - \tilde{P}_{j-1})\tilde{\Theta} g\|_{L_{p'}(\mathbb{R}^d)})^{q'}\right)^{\frac{1}{q'}} \lesssim \|\tilde{\Theta} g\|_{L_{p'}(\mathbb{R}^d)},$$

since we assumed that the sequence $\{\tilde{P}_j\}_{j\geq 0}$ is uniformly bounded in $L_{p'}$. We continue by introducing the operator Θ in analogy to $\tilde{\Theta}$

$$\Theta : g \longmapsto \sum_{l\geq 0} \rho^{-sl}(P_l - P_{l-1})g. \tag{6.1.14}$$

Then we bound $\|\tilde{\Theta} g\|_{L_{p'}(\mathbb{R}^d)}$ by

$$\begin{aligned}
\|\tilde{\Theta} g\|_{L_{p'}(\mathbb{R}^d)} &= \sup_{f\in L_p(\mathbb{R}^d), f\neq 0} \frac{|\langle \tilde{\Theta} g, f\rangle|}{\|f\|_{L_p(\mathbb{R}^d)}} \\
&= \sup_{f\in L_p(\mathbb{R}^d), f\neq 0} \frac{|\langle g, \Theta f\rangle|}{\|f\|_{L_p(\mathbb{R}^d)}} \\
&\leq \sup_{f\in L_p(\mathbb{R}^d), f\neq 0} \frac{\|g\|_{(B^s_q(L_p(\mathbb{R}^d)))'}\|\Theta f\|_{B^s_q(L_p(\mathbb{R}^d))}}{\|f\|_{L_p(\mathbb{R}^d)}}.
\end{aligned}$$

If we now show that $\frac{\|\Theta f\|_{B^s_q(L_p(\mathbb{R}^d))}}{\|f\|_{L_p(\mathbb{R}^d)}}$ is bounded we are done. To this end we use the upper inequality of (5.1.10) to write

$$\begin{aligned}
\|\Theta f\|_{B^s_q(L_p(\mathbb{R}^d))} &\lesssim \|\Theta f\|_{(s,\infty,p,q)} \\
&= \left(\sum_{j=0}^{m}(\rho^{sj}\|\rho^{-sj}(P_j - P_{j-1})f\|_{L_p(\mathbb{R}^d)})^q\right)^{1/q} \\
&= \left(\sum_{j=0}^{m}\|(P_j - P_{j-1})f\|^q_{L_p(\mathbb{R}^d)}\right)^{1/q} \\
&\lesssim \|f\|_{L_p(\mathbb{R}^d)},
\end{aligned}$$

where we used the identity (6.1.13) in the second line and a stability criterion in the last inequality. □

Remark 6.1.3. *In Triebel (1983, Sec. 2.11) it was shown that for* $1 \le p < \infty$*,* $0 < q < \infty$ *and* $s > 0$ *one has*

$$(B_q^s(L_p(\mathbb{R}^d)))' = B_{q'}^{-s}(L_{p'}(\mathbb{R}^d)).$$

In particular, we are thus able to characterize Besov *spaces with negative smoothness index which is specified in the following corollary.*

Corollary 6.1.2. *Let the assumptions of Theorem 6.1.1 hold. Further, let* $-1 + \frac{1}{p} < s < \frac{1}{p}$*. Then for any function* $f \in B_q^s(L_p(\mathbb{R}^d))$ *one has*

$$\|f\|_{B_{q'}^{-s}(L_{p'}(\mathbb{R}^d))} \sim \|f\|^{\sim}_{(-s,\infty,p,q)}. \tag{6.1.15}$$

In the same manner as we have characterized the primal Besov space by the dual wavelet coefficients in Section 5.2 this also works if we exchange the primal and the dual basis. With this assumption we arrive at the following norm equivalence for Besov spaces with negative smoothness index. Here we mimic the proof of Cohen (2003, Sec. 3.8).

Theorem 6.1.3. *Assume that for* $f \in L_p(\mathbb{R}^d)$*,* $1 \le p, q \le \infty$*,* $s > 0$ *we have for the dual projectors* $\tilde{P}_j$ *and* $\tilde{Q}_j$ *(6.1.4) and (6.1.5) the characterization*

$$\|f\|_{B_q^s(L_p(\mathbb{R}^d))} \sim \|\tilde{P}_0 f\|_{L_p(\mathbb{R}^d)} + \left(\sum_{j=0}^{\infty} (\rho^{js} \|\tilde{Q}_j f\|_{L_p(\mathbb{R}^d)})^q \right)^{\frac{1}{q}}. \tag{6.1.16}$$

Then it follows the dual characterization

$$\|f\|_{B_{q'}^{-s}(L_{p'}(\mathbb{R}^d))} \sim \|P_0 f\|_{L_{p'}(\mathbb{R}^d)} + \left(\sum_{j=0}^{\infty} (\rho^{-js} \|Q_j f\|_{L_{p'}(\mathbb{R}^d)})^{q'} \right)^{\frac{1}{q'}} \tag{6.1.17}$$

for $\frac{1}{p} + \frac{1}{p'} = 1$ *and* $\frac{1}{q} + \frac{1}{q'} = 1$*.*

Proof. Here we choose the dual norm of the form

$$\|f\|_{B_{q'}^{-s}(L_{p'}(\mathbb{R}^d))} = \sup_{\substack{g \in \mathcal{D} \\ \|g\|_{B_q^s(L_p(\mathbb{R}^d))} = 1}} \langle f, g \rangle, \tag{6.1.18}$$

where $\mathcal{D}$ is the space of infinitely often differentiable functions with compact support. We mean by $\langle f, g \rangle$ the duality in the sense of $(\mathcal{D}, \mathcal{D}')$ with $\mathcal{D}'$ denoting the space of distributions. Note that this is equivalent to the definition (6.1.10) since the ball with $\|g\|_{B_q^s(L_p(\mathbb{R}^d))} = 1$ is symmetric.

If the function f is also a distribution it follows

$$\begin{aligned}
\langle f,g\rangle &= \langle P_0 f + \sum_{j\geq 0} Q_j f, \tilde{P}_0 g + \sum_{j\geq 0}\tilde{Q}_j g\rangle \\
&= \langle P_0 f, \tilde{P}_0 g\rangle + \sum_{j\geq 0}\langle Q_j f, \tilde{Q}_j g\rangle \\
&\leq \|P_0 f\|_{L_{p'}(\mathbb{R}^d)}\|\tilde{P}_0 g\|_{L_p(\mathbb{R}^d)} + \sum_{j\geq 0}\|Q_j f\|_{L_{p'}(\mathbb{R}^d)}\|\tilde{Q}_j g\|_{L_p(\mathbb{R}^d)} \\
&\lesssim \|g\|_{B^s_q(L_p(\mathbb{R}^d))}\left(\|P_0 f\|_{L_{p'}(\mathbb{R}^d)} + \left(\sum_{j=0}^{\infty}(\rho^{-js}\|Q_j f\|_{L_{p'}(\mathbb{R}^d)})^{q'}\right)^{\frac{1}{q'}}\right),
\end{aligned}$$

where we first used a continuous HÖLDER inequality and then a discrete one in the third line. The fact that we can write f and g as a direct sum due to $V_0 \perp \tilde{W}_j$ and also $\tilde{V}_0 \perp W_j$ for $j \geq 0$ and the biorthogonality relation $W_j \perp \tilde{W}_k = \{0\}$ if $j \neq k$ shows the last inequality. Since distributions are dense in $B^{-s}_{q'}(L_{p'}(\mathbb{R}^d))$ we hence have proved the upper inequality of (6.1.17).

Now we turn to the lower inequality. First we use the definition of the duality for discrete sequence norms in ℓ_q and $\ell_{q'}$. Thus, we find a sequence $a = (a_j)_{j\geq 0}$ with $\|a\|_{\ell_q} = 1$ such that

$$\begin{aligned}
\|P_0 f\|_{L_{p'}(\mathbb{R}^d)} &+ \left(\sum_{j=0}^{\infty}(\rho^{-js}\|Q_j f\|_{L_{p'}(\mathbb{R}^d)})^{q'}\right)^{\frac{1}{q'}} \\
&= \|P_0 f\|_{L_{p'}(\mathbb{R}^d)} + \sum_{j=0}^{\infty} a_j \rho^{-js}\|Q_j f\|_{L_{p'}(\mathbb{R}^d)}. \qquad (6.1.19)
\end{aligned}$$

Now we rewrite $\|Q_j f\|_{L'_p(\mathbb{R}^d)}$ and $\|P_0 f\|_{L'_p(\mathbb{R}^d)}$. To do that we first employ the norm equivalences

$$\|\sum_{\lambda\in\nabla_j} c_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} \sim \|\sum_{\lambda\in\nabla_j} d_\lambda\tilde{\psi}_\lambda\|_{L_p(\mathbb{R}^d)} \sim |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\|(c_\lambda)_{\lambda\in\nabla_j}\|_{\ell_p} \qquad (6.1.20)$$

and the identity

$$\langle \sum_{\lambda\in\nabla_j} c_\lambda\psi_\lambda, \sum_{\lambda\in\nabla_j} d_\lambda\tilde{\psi}_\lambda\rangle = \sum_{\lambda\in\nabla_j} c_\lambda\bar{d}_\lambda, \qquad (6.1.21)$$

(which is a direct consequence of the biorthogonality of the functions ϕ and $\tilde{\phi}$). They imply that there exist functions $h_j \in \tilde{W}_j$ such that $\|h_j\|_{L_p(\mathbb{R}^d)} = 1$ such that we can rewrite $\|Q_j f\|_{L'_p(\mathbb{R}^d)}$ and $\|P_0 f\|_{L'_p(\mathbb{R}^d)}$ as

$$\|Q_j f\|_{L'_p(\mathbb{R}^d)} = \sup_{h \in \tilde{W}_j, \|h\|_{L_p(\mathbb{R}^d)}=1} \langle Q_j f, h \rangle. \tag{6.1.22}$$

where we define $h_j \in \tilde{W}_j$ to be the function for that this equality holds. Then we get clearly

$$\|Q_j f\|_{L'_p(\mathbb{R}^d)} \lesssim \langle Q_j f, h_j \rangle. \tag{6.1.23}$$

Similarly we find a function $g_0 \in \tilde{V}_0$ with $\|g_0\|_{L_p(\mathbb{R}^d)} = 1$ such that

$$\|P_0 f\|_{L'_p(\mathbb{R}^d)} \leq \langle P_0 f, g_0 \rangle. \tag{6.1.24}$$

We set

$$g := g_0 + \sum_{j \geq 0} a_j \rho^{-js} h_j. \tag{6.1.25}$$

Again, employing biorthogonality and the decomposition in a direct sum we get by the above estimates of $\|Q_j f\|_{L'_p(\mathbb{R}^d)}$ and $\|P_0 f\|_{L'_p(\mathbb{R}^d)}$

$$\begin{aligned}
\langle f, g \rangle &= \langle P_0 f + \sum_{j \geq 0} Q_j f, g_0 + \sum_{j \geq 0} a_j \rho^{-js} h_j \rangle && (6.1.26)\\
&= \langle P_0 f, g_0 \rangle + \sum_{j \geq 0} a_j \rho^{-js} \langle Q_j f, h_j \rangle && (6.1.27)\\
&\gtrsim \|P_0 f\|_{L'_p(\mathbb{R}^d)} + \sum_{j \geq 0} a_j \rho^{-js} \|Q_j f\|_{L'_p(\mathbb{R}^d)}. && (6.1.28)
\end{aligned}$$

By assumption we have a characterization of $B^s_q(L_p(\mathbb{R}^d))$ by (6.1.16). Together with the fact $\|g_0\|_{L_p(\mathbb{R}^d)} = \|h_j\|_{L_p(\mathbb{R}^d)} = 1$ we get the following bound for $\|g\|_{B^s_q(L_p(\mathbb{R}^d))}$

$$\begin{aligned}
\|g\|_{B^s_q(L_p(\mathbb{R}^d))} &\lesssim \|\tilde{P}_0 g\|_{L_p(\mathbb{R}^d)} + \left(\sum_{j=0}^{\infty} (\rho^{js} \|\tilde{Q}_j g\|_{L_p(\mathbb{R}^d)})^q \right)^{\frac{1}{q}} \\
&= \|g_0\|_{L_p(\mathbb{R}^d)} + \left(\sum_{j=0}^{\infty} (a_j \|h_j\|_{L_p(\mathbb{R}^d)})^q \right)^{\frac{1}{q}} = 2.
\end{aligned}$$

Now we build the supremum in (6.1.26) over the ball $\|g\|_{B^s_q(L_p(\mathbb{R}^d))} \lesssim 2$ (which will increase the constant) to establish the lower inequality of (6.1.17). This arises in

$$\|P_0 f\|_{L_{p'}(\mathbb{R}^d)} + \left(\sum_{j=0}^{\infty} (\rho^{-js} \|Q_j f\|_{L_{p'}(\mathbb{R}^d)})^{q'} \right)^{\frac{1}{q'}} \leq B^{-s}_{q'}(L_{p'}(\mathbb{R}^d)), \quad (6.1.29)$$

which concludes the proof. □

Chapter 7

Unstable Approximation

In the last chapter we have seen that Besov spaces can be characterized by means of the projectors P_j with L_p norms in the scale $1 \le p \le \infty$. We have seen that in this setting these projectors are uniformly bounded which was a fundamental condition to prove the norm equivalences in the last chapter.

Actually, Besov spaces can be defined also for the range $0 < p < 1$. This restriction is due to the fact that the projectors $P_j f$ are not bounded in the L_p quasi-norm. In particular, the L_p boundedness of a function f does not even imply that it is a distribution, i.e. the evaluation of inner products $\langle f, \tilde{\phi}_{j,k} \rangle$ can be meaningless.

In the chapter for nonlinear approximation we will show that the range $0 < p < 1$ has an eminent impact.

7.1 Stability, Jackson and Bernstein Type Inequalities

We focus now on the range of L_p spaces with $0 < p < 1$. Despite that the projections $P_j f$ are not bounded in L_p we nevertheless prove a stability result for the generator functions ϕ.

To prove this stability we cannot use the same technique as for $p \ge 1$. Here we have to put further assumptions on the generator function ϕ. In particular we assume that the generator functions $\phi_{j,k}$ are *locally linear independent*. A function f is called *locally linear independent* if $\sum_{k \in \mathbb{Z}^d} c_k f(\cdot - k)$ vanishes on some domain J, then it follows that $c_k = 0$ whenever $|\operatorname{supp} f(\cdot - k) \cap J| \ne 0$. It was proved that generator functions fulfill this property in the univariate setting for any non-trivial interval $[a, b]$ in Lemari-Rieusset and Malgouyres (1991). A proof for multivariate separable scaling functions is given in Cohen (2003). In the univariate case the support of the refinable functions is always a closed interval. It is shown that the support equals to the attractor of an iterated function system under the assumption of local linear independence (Cheung, Tang, and Zhou (2002)). The local linear independence of refinable vectors of functions was treated in Goodman, Jia, and Zhou (2000) and Sun (2001). A complete characterization for the local linear independence of multivariate refinable functions is given by finite matrix products, strictly in terms of the refinement mask (Presentation on Canada-China Math Congress 2001, Vancouver/BC by Hoi Ling Cheung (Hong Kong)). An adaptation of the proof of Cohen for the non-separable setting needs the distinction of some more cases but can be done straightforward using the same idea.

As we already mentioned we cannot hope to have the projectors P_j bounded in $L_p(\mathbb{R}^d)$ for $p < 1$. Nonetheless we can show a stability result.

Theorem 7.1.1 (Stability). *Let* $(\phi, \tilde{\phi})$ *be a biorthogonal pair of* $(\mathbf{h}, M)$*-refinable (or* $(\tilde{\mathbf{h}}, M)$*-refinable, resp.) scaling functions in the sense of (2.1.19). Assume that* $\phi \in L_p(\mathbb{R}^d)$ *for* $0 < p < 1$*. Further, we assume that the functions* $\phi_{j,k}$ *are locally linear independent. Then*

$$\left\| \sum_{k \in \mathbb{Z}^d} c_k \phi_{j,k} \right\|_{L_p(\mathbb{R}^d)} \sim |\det M|^{j(\frac{1}{2}-\frac{1}{p})} \|(c_k)_{k \in \mathbb{Z}^d}\|_{\ell_p} \tag{7.1.1}$$

independently of j.

Proof. Again we first consider the inequality on the cubes $I_{j,k} := \|M^{-j}\|_\infty (k + [0,1]^d)$, $k \in \mathbb{Z}^d$. Since the support of ϕ is compact we obtain the upper inequality

$$\begin{aligned} \left\| \sum_{k \in \mathbb{Z}^d} c_k \phi_{j,k} \right\|_{L_p(I_{j,l})} &\lesssim \sup_{k:\ l-k \,\in\, \operatorname{supp} \phi} |c_k| \cdot \|\phi_{j,0}\|_{L_p(\mathbb{R}^d)} \\ &\lesssim |\det M|^{j(\frac{1}{2}-\frac{1}{p})} \|(c_k)_{k \in \mathbb{Z}^d}\|_{\ell_p} \end{aligned}$$

where we have used that $\|\cdot\|_{\ell_\infty} \sim \|\cdot\|_{\ell_p}$ on finite index sets.

Assume now that there exists a cube $J \subset \mathbb{R}^d$ such that the local linear independence holds on J at level $j = 0$. Defining $J_{j,k} := \|M^{-j}\|_\infty(k + J)$ we obtain that

$$\left\| \sum_{k \in \mathbb{Z}^d} c_k \phi_{j,k} \right\|_{L_p(J_{j,l})} \sim |\det M|^{j(\frac{1}{2}-\frac{1}{p})} \|(c_k)_{k \in \{k:\ l-k \in \operatorname{supp}\phi - J\}}\|_{\ell_p}$$

using again that all quasi-norms are equivalent on finite index sets. Raising to the power p and summation over $l \in \mathbb{Z}^d$ proves the global stability. □

Theorem 7.1.2 (Jackson Inequality). *Let $\phi \in L_p(\mathbb{R}^d)$ for $p > 0$ be a $(\mathbf{h}, M)$-refinable function. Furthermore let $n-1$ be the order of polynomial reproduction of ϕ in S_j. Then we get the* Jackson *type inequality*

$$\inf_{g \in V_j} \|f - g\|_{L_p(\mathbb{R}^d)} \lesssim \omega_n(f, \rho^{-j})_p, \tag{7.1.2}$$

where ρ denotes the spectral norm of the scaling matrix M.

Proof. We consider the cubes $I_{j,k} := \rho^{-j}(k + [0,1]^d)$. Now we introduce two sets

$$E_l := \{k \in \mathbb{Z}^d :\ |\operatorname{supp}\phi_{j,l} \cap I_{j,k}| \neq 0\}$$

and

$$F_k := \{l \in \mathbb{Z}^d :\ |\operatorname{supp}\phi_{j,l} \cap I_{j,k}| \neq 0\}.$$

There exists a constant $C > 0$ such that $|E_l| = C \cdot |F_k|$ independently of j and k. Now we define larger sets

$$\tilde{J}_{j,k} := \bigcup_{l \in F_k} \operatorname{supp}\phi_{j,l}.$$

For each of these sets we choose a polynomial $p_{j,k} \in \Pi_{n-1}$ such that

$$\|f - p_{j,k}\|_{L_p(\tilde{J}_{j,k})} \le 2 \inf_{g \in \Pi_{n-1}} \|f - g\|_{L_p(\tilde{J}_{j,k})}. \tag{7.1.3}$$

As we have seen in Section 3.3, Lemma 3.3.2 we apply now a Whitney type estimate of the form (3.3.7). It were Oswald and Storozhenko (1978) who extended this estimate to the case $0 < p < 1$. With the help of this we are able to obtain

$$\|f - p_{j,k}\|_{L_p(\tilde{J}_{j,k})} \lesssim \omega_n(f, \rho^{-j})_{L_p(\tilde{J}_{j,k})}. \tag{7.1.4}$$

We can uniquely expand each of these polynomials into a representation in terms of the scaling function basis

$$p_{j,k} = \sum_{l \in \mathbb{Z}^d} b_{kjl} \phi_{j,l}. \tag{7.1.5}$$

Now we approximate the function f in V_j by

$$f_j = \sum_{l \in \mathbb{Z}^d} \tilde{c}_{j,l} \phi_{j,l}. \tag{7.1.6}$$

with the particular choice

$$\tilde{c}_{j,l} = \frac{1}{|E_l|} \sum_{k \in E_l} b_{kjl}. \tag{7.1.7}$$

This can be seen as a smoothing of the coefficients of the local polynomial approximation. To estimate the error of the local polynomial approximation we use the WHITNEY estimate (7.1.4), the expansion of $p_{j,k}$ (7.1.5) and the expansion of f (7.1.6) with the coefficients (7.1.7). Together with the stability result (7.1.1) we derive

$$\begin{aligned} \|f - f_j\|_{L_p(I_{j,k})} &\lesssim \|f - p_{j,k}\|_{L_p(I_{j,k})} + \|f_j - p_{j,k}\|_{L_p(I_{j,k})} \\ &\lesssim \omega_n(f, \rho^{-j})_{L_p(\tilde{J}_{j,k})} + \left(\sum_{l \in F_k} |b_{kjl} - \tilde{c}_{j,l}|^p \right)^{1/p} . \end{aligned}$$

To estimate the second term we use the expansion of the coefficients $\tilde{c}_{j,l}$ in (7.1.7) and apply the triangle inequality, i.e.

$$\begin{aligned} \left(\sum_{l \in F_k} |b_{kjl} - \tilde{c}_{j,l}|^p \right)^{1/p} &\lesssim \left(\sum_{l \in F_k} |b_{kjl} - \frac{1}{|E_l|} \sum_{k \in E_l} b_{kjl}|^p \right)^{1/p} \\ &\lesssim \left(\sum_{l \in F_k} \sum_{m \in E_l} |b_{kjl} - b_{mjl}|^p \right)^{1/p} . \end{aligned}$$

Altogether we get

$$\|f - f_j\|_{L_p(I_{j,k})} \lesssim \omega_n(f, \rho^{-j})_{L_p(\tilde{J}_{j,k})} + \left(\sum_{l \in F_k} \sum_{m \in E_l} |b_{kjl} - b_{mjl}|^p \right)^{1/p} .$$

For the estimation of the second term we employ the local linear independence of polynomials that is known to be valid on any domain of non-zero measure which implies

$$\begin{aligned}\sum_{m\in E_l}|b_{kjl}-b_{mjl}|^p &\lesssim \sum_{m\in E_l}\|p_{j,k}-p_{j,m}\|^p_{L_p(\operatorname{supp}\phi_{j,l})}\\ &\lesssim \sum_{m\in E_l}\|f-p_{j,m}\|^p_{L_p(\operatorname{supp}\phi_{j,l})}\\ &\lesssim \sum_{m\in E_l}\|f-p_{j,m}\|^p_{L_p(\tilde J_{j,m})}\\ &\lesssim \sum_{m\in E_l}\omega_n(f,\rho^{-j})^p_{L_p(\tilde J_{j,m})},\end{aligned}$$

where we have used again (7.1.4) in the last inequality. Combining this we obtain

$$\sum_{l\in F_k}\sum_{m\in E_l}|b_{kjl}-b_{mjl}|^p \lesssim \sum_{l\in F_k}\sum_{m\in E_l}\omega_n(f,\rho^{-j})^p_{L_p(\tilde J_{j,m})}.$$

Taking the p-th power of $\|f-f_j\|_{L_p(I_{j,k})}$ implicates

$$\|f-f_j\|^p_{L_p(I_{j,k})} \lesssim \omega_n(f,\rho^{-j})^p_{L_p(I_{j,k})} + \sum_{l\in F_k}\sum_{m\in E_l}\omega_n(f,\rho^{-j})^p_{L_p(\tilde J_{j,m})}.$$

In order to turn this into a global estimate we encounter a difficulty since it is not clear that

$$\sum_{k\in\mathbb{Z}^d}\sum_{l\in F_k}\sum_{m\in E_l}\omega_n(f,\rho^{-j})^p_{L_p(\tilde J_{j,m})} \lesssim \omega_n(f,\rho^{-j})^p_{L_p(\mathbb{R}^d)}.$$

For this reason we introduce the following variant of the modulus of smoothness: the *averaged modulus of smoothness*

$$\tilde\omega_n(f,t,\Omega)_p := \left(\frac{1}{t^d}\int_{[0,t]^d}\|\Delta_h^n f\|^p_{L_p(\Omega_{h,n})}\mathrm{d}h\right)^{1/p} \tag{7.1.8}$$

for $p<\infty$ and $\Omega_{h,n}$ as defined in (2.5.20).

We easily estimate $\tilde\omega_n(f,t,\Omega)_p \lesssim 2\omega_n(f,t,\Omega)_p$. It is more challenging to show the inverse direction. Finally one has the full equivalence assuming that Ω is a cube and there exists a constant $C>0$ such that $t<C\cdot\operatorname{diam}(\Omega)$

$$\tilde\omega_n(f,t,\Omega)_p \sim \omega_n(f,t,\Omega)_p, \tag{7.1.9}$$

where the constant here only depends on C, p and n. A proof for the univariate case can be found in DeVore and Lorentz (1993, p.185). An adaptation to the multivariate case is shown in Cohen (2003, p.169).

A nice property of the averaged modulus of smoothness is that it satisfies

$$\sum_{k\in\mathbb{Z}^d}\sum_{l\in F_k}\sum_{m\in E_l}\tilde{\omega}_n(f,\rho^{-j},\tilde{J}_{j,m})_p^p \lesssim \tilde{\omega}_n(f,\rho^{-j})_p^p.$$

Now we are ready to conclude the proof by summing $\|f-f_j\|_{L_p(I_{j,k})}$ over k and employ the equivalence (7.1.9) to show

$$\|f-f_j\|_{L_p(\mathbb{R}^d)} \lesssim \omega_n(f,\rho^{-j})_p.$$

□

Theorem 7.1.3 (Bernstein Inequality). *If $\phi \in L_p(\mathbb{R}^d)$ for $p > 0$ and if $\phi \in B_q^n(L_p(\mathbb{R}^d))$, $n \in \mathbb{N}$ for some $q > 0$ one has for $f \in V_j$*

$$\omega_n(f,t)_p \lesssim \left(\min\{1, t\rho^j\}\right)^n \|f\|_{L_p(\mathbb{R}^d)}. \tag{7.1.10}$$

Proof. For the representation $f = \sum_{k\in\mathbb{Z}^d} c_k\phi_{j,k}$ we obtain as in the proof of the Bernstein inequality in Theorem 4.1.3

$$\omega_n(f,t)_p \lesssim \omega_n(\phi_{j,0},t)_p\|(c_k)_{k\in\mathbb{Z}^d}\|_{\ell_p}.$$

Again, by monotonicity of the modulus it is sufficient to show the inequality for $t \le \rho^{-j}$. Therefore it is enough to consider $t = \rho^{-l}$ for $l > j$. It follows

$$\omega_n(\phi_{j,0},\rho^{-l})_p \le |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\omega_n(\phi,\rho^{j-l})_p,$$

where we have used the scaling property (2.5.29) of the modulus of smoothness. By assumption $\phi \in B_q^n(L_p(\mathbb{R}^d))$ one derives that $\omega_n(\phi,\rho^{j-l})_p$ decays at least like $2^{n(j-l)}$. Since $l > j$ we can bound $\omega_n(\phi,\rho^{j-l})_p$ by one.

Combination of this result and the stability property for $p < 1$ proves the desired Bernstein inequality. □

Another version of a Bernstein estimate involves Besov norms.

Theorem 7.1.4. *Assume that ϕ is $(\mathbf{a}, M)$-refinable function in $B_q^s(L_p(\mathbb{R}^d))$ for $q, s > 0$. Then one has for $f \in V_j$*

$$\|f\|_{B_q^s(L_p(\mathbb{R}^d))} \lesssim \rho^{sj}\|f\|_{L_p(\mathbb{R}^d)}. \tag{7.1.11}$$

Proof. For $f = \sum_{k\in\mathbb{Z}^d} c_k\phi_{j,k} \in V_j$ we use the properties of the modulus of smoothness to obtain again

$$\omega_n(f,t)_p \lesssim \omega_n(\phi_{j,0},t)_p \|(c_k)_{k\in\mathbb{Z}^d}\|_{\ell_p}.$$

This implies

$$|f|_{B^s_q(L_p(\mathbb{R}^d))} \lesssim \|(c_k)_{k\in\mathbb{Z}^d}\|_{\ell_p} |\phi_{j,0}|_{B^s_q(L_p(\mathbb{R}^d))}.$$

It remains to evaluate the seminorm $|\phi_{j,0}|_{B^s_q(L_p(\mathbb{R}^d))}$. For the same reasons as in the proof of equation (7.1.10) we use a monotonicity argument for the modulus of smoothness. Then we have

$$\begin{aligned}\omega_n(\phi_{j,0},\rho^{-l})_p &\leq |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\omega_n(\phi,\rho^{j-l})_p \\ &\lesssim |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\rho^{j-l}\epsilon^{\phi}_{l-j}\end{aligned}$$

where $(\epsilon^{\phi}_n)_{n\geq 0}$ is an ℓ_q sequence with $\|(\epsilon^{\phi}_n)_{n\geq 0}\|_{\ell_q} \leq |\phi|_{B^s_q(L_p(\mathbb{R}^d))}$.

When $l \leq j$ we use the rough estimate

$$\omega_n(\phi_{j,0},\rho^{-l})_p \lesssim \|\phi_{j,0}\|_{L_p(\mathbb{R}^d)} \lesssim |\det M|^{j(\frac{1}{2}-\frac{1}{p})}.$$

Employing these two estimates we are now able to bound $|\phi_{j,0}|_{B^s_q(L_p(\mathbb{R}^d))}$.

$$\begin{aligned}|\phi_{j,0}|_{B^s_q(L_p(\mathbb{R}^d))} &\sim \|\left(\rho^{sl}\omega_n(\phi_{j,0},\rho^{-l})_p\right)_{l\geq 0}\|_{\ell_q} \\ &\leq \|\left(\rho^{sl}\omega_n(\phi_{j,0},\rho^{-l})_p\right)_{0\leq l\leq j}\|_{\ell_q} \\ &\quad +\|\left(\rho^{sl}\omega_n(\phi_{j,0},\rho^{-l})_p\right)_{l>j}\|_{\ell_q} \\ &\leq |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\left(C+\|\left(\rho^{sl}\rho^{s(j-l)}\epsilon^{\phi}_{l-j}\right)_{l>j}\|_{\ell_q}\right) \\ &\lesssim |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\rho^{sj},\end{aligned}$$

where $C > 0$ is a constant. Now we use the lower inequality oft the stability relation (7.1.1) for

$$\|f\|_{B^s_q(L_p(\mathbb{R}^d))} \lesssim |\det M|^{j(\frac{1}{2}-\frac{1}{p})}\|(c_k)_{k\in\mathbb{Z}^d}\|_{\ell_p}\rho^{sj}$$

and conclude the proof. □

7.2 Norm Equivalences

Now we are ready to state the main result of this chapter. We extend the characterization $\mathcal{A}_q^s(L_p(\mathbb{R}^d)) = B_q^\alpha(L_p(\mathbb{R}^d))$ of Theorem 5.2.1 in Section 5.2 to the range $0 < p, q < 1$.

Theorem 7.2.1. *Let $(\phi, \tilde{\phi})$ be a biorthogonal pair of $(\mathbf{h}, M)$-refinable (or $(\tilde{\mathbf{h}}, M)$-refinable, resp.) scaling functions in the sense of (2.1.19). Assume that $\phi \in L_p(\mathbb{R}^d)$ and $\tilde{\phi} \in L_{p'}(\mathbb{R}^d)$ such that $\frac{1}{p} + \frac{1}{p'} = 1$. If $\phi \in B_q^\alpha(L_p(\mathbb{R}^d))$ for some $q > 0$, then we have for all $s < \min\{n, \alpha\}$ where $n-1$ is the order of polynomial reproduction of the spaces V_j,*

$$\|f\|_{\mathcal{A}_q^s(L_p(\mathbb{R}^d))} \sim \|f\|_{B_q^s(L_p(\mathbb{R}^d))}. \tag{7.2.1}$$

Proof. We tackle the proof by comparison of $\operatorname{dist}_p(f, V_j)$ and $\omega_n(f, \rho^{-j})_p$.

The lower inequality directly follows by the Jackson type inequality (7.1.2).
To show the converse inequality we use the Bernstein type inequality (7.1.11) which implies the simpler inverse estimate

$$\omega_n(f,t)_p \lesssim \left(\min\{1, t\rho^j\}\right)^\alpha \|f\|_{L_p(\mathbb{R}^d)}, \tag{7.2.2}$$

if $f \in V_j$.

Again, we only consider $\omega_n(f,t)_p$ for $t = \rho^{-l}$ with $l > j$ since this estimate is trivial for $t \geq \rho^{-j}$ due to the monotonicity of the modulus of smoothness.

For $f \in L_p(\mathbb{R}^d)$ we let $f_j \in V_j$ such that

$$\|f - f_j\|_{L_p(\mathbb{R}^d)} \leq 2 \operatorname{dist}_p(f, V_j). \tag{7.2.3}$$

We expand f_j by

$$f_j = \sum_{m=0}^{j} f_m \qquad \text{with } f_m \in V_m.$$

Thus by application of the p triangle inequality

$$\|f+g\|_{L_p(\mathbb{R}^d)}^p \leq \|f\|_{L_p(\mathbb{R}^d)}^p + \|g\|_{L_p(\mathbb{R}^d)}^p \tag{7.2.4}$$

we obtain

$$\begin{aligned}
\omega_n(f,\rho^{-j})_p^p &\le \omega_n(f-f_j,\rho^{-j})_p^p+\omega_n(f_j,\rho^{-j})_p^p && (7.2.5)\\
&\lesssim \omega_n(f_0,\rho^{-j})_p^p+\sum_{l=0}^{j-1}\omega_n(f_{l+1}-f_l,\rho^{-j})_p^p && (7.2.6)\\
&\quad +\omega_n(f-f_j,\rho^{-j})_p^p\\
&\lesssim \rho^{-sj}\|f_0\|_{L_p(\mathbb{R}^d)}^p+\rho^{-sj}\sum_{l=0}^{j-1}\rho^{sl}\|f_{l+1}-f_l\|_{L_p(\mathbb{R}^d)}^p && (7.2.7)\\
&\quad +\|f-f_j\|_{L_p(\mathbb{R}^d)}^p\\
&\lesssim \rho^{-sj}\|f_0\|_{L_p(\mathbb{R}^d)}^p+\rho^{-sj}\sum_{l=0}^{j}\rho^{sl}\|f-f_l\|_{L_p(\mathbb{R}^d)}^p. && (7.2.8)
\end{aligned}$$

We have used from (7.2.6) to (7.2.7) the simpler BERNSTEIN type estimate (7.2.2) and the fact that $f_{l+1}-f_l\in V_{l+1}$. We also get with the triangle inequality

$$\begin{aligned}
\|f_0\|_{L_p(\mathbb{R}^d)}^p &\lesssim \|f-f_0\|_{L_p(\mathbb{R}^d)}^p+\|f\|_{L_p(\mathbb{R}^d)}^p\\
&\le 2\,\mathrm{dist}_p(f,V_0)^p+\|f\|_{L_p(\mathbb{R}^d)}^p\le 3\|f\|_{L_p(\mathbb{R}^d)}^p.
\end{aligned}$$

This finally yields

$$\omega_n(f,\rho^{-j})_p^p\lesssim\rho^{-sj}\|f\|_{L_p(\mathbb{R}^d)}^p+\rho^{-sj}\sum_{l=0}^{j}\rho^{sl}\,\mathrm{dist}_p(f,V_l)^p. \qquad (7.2.9)$$

For the first part of this inequality we deduce with $s>t$

$$\left(\sum_{j\ge0}(\rho^{jt}\rho^{-js}\|f\|_{L_p(\mathbb{R}^d)}^p)^q\right)\lesssim\|f\|_{L_p(\mathbb{R}^d)}^p. \qquad (7.2.10)$$

For the second part we utilize again the discrete Hardy inequality (5.1.8) with

$$\begin{aligned}
a_j &:= \mathrm{dist}_p(f,V_j)\\
b_j &:= \rho^{-js}\sum_{l=0}^{j}\rho^{sl}a_l,
\end{aligned}$$

which finally shows that

$$\|f\|_{B^s_q(L_p(\mathbb{R}^d))} \lesssim \|f\|_{\mathcal{A}^s_q(L_p(\mathbb{R}^d))}. \tag{7.2.11}$$

□

Chapter 8

Nonlinear Approximation

8.1 Introduction

In Theorem 5.2.1 we have seen the characterization

$$f \in \mathcal{A}_q^\alpha(L_p(\mathbb{R}^d)) \Longleftrightarrow f \in B_q^\alpha(L_p(\mathbb{R}^d)) \tag{8.1.1}$$

for non-separable wavelet decompositions with isotropic scaling matrices (cf. Figure 5.1). We have also seen that examining the coefficients of the wavelet expansions determines when a function belongs to this Besov space. This approximation result expresses that approximation order $\mathcal{O}(\rho^{-j\alpha})$ in $L_p(\mathbb{R}^d)$ as the refinement or discretization tends *uniformly* to zero always implies smoothness order α in $L_p(\mathbb{R}^d)$. At the same time we cannot expect that we get a high approximation order if the function has low smoothness in $L_p(\mathbb{R}^d)$. We recall that the approximation is described by $\mathrm{dist}_p(f, V_j) \lesssim \rho^{-j\alpha}$. If the smoothness α is small then ρ^j should be large to achieve a small approximation order. That is one has to incorporate wavelet coefficients corresponding to small details and thus a high number of coefficients.

The most popular applications of wavelet decompositions as discretization for numerical schemes are wavelet-Galerkin schemes to solve partial differential equations or compression of images. The heart of those algorithms is concerned with *nonlinear approximation*.

Here we remark that a frequent phenomenon is the occurrence of singularities in the approximated function. These singularities should not be seen as non-relevant parts of this function. Quite contrary to that they often have important physical relevance. That's why is should be natural to use an *adaptive* refinement procedure to resolve even less smooth parts accurately. Adaptivity is understand here in the sense that the refinement is allowed to be local. Around non-smooth parts of the underlying function the refinement is finer whereas larger building blocks can be used for smooth parts.

8.2 N-Term Wavelet Approximation

Assume that a function f has the wavelet expansion

$$f = \sum_{e\in E} \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^d} d_{e,j,k}\psi_{e,j,k}. \tag{8.2.1}$$

For numerical applications we must replace this sum by a finite set. The question is how to manage this in the most efficient way. For this purpose we introduce a set of nonlinear approximation sets that play the analogous role for the shift-invariant spaces V_j. We employ here and in what follows again the notations from Section 2.2. We want to approximate the function f by an element from the set

$$\Sigma_n := \left\{ S = \sum_{\lambda\in\Lambda\subset\nabla} d_\lambda \psi_\lambda : \ |\Lambda| \le n \right\}. \tag{8.2.2}$$

That is, we choose n suitable wavelet coefficients to approximate a function and minimize the approximation error, see Figure 8.1 for visualization.

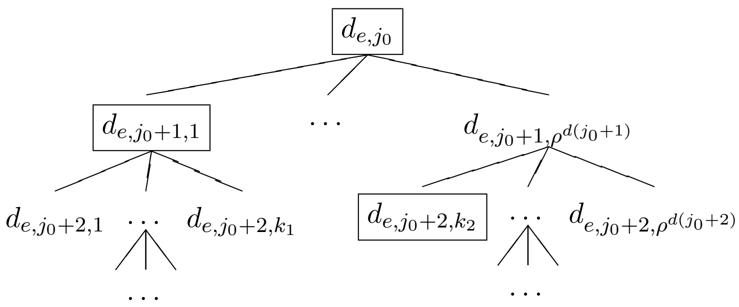


Figure 8.1: This graph may represent one adaptive choice of n suitable wavelet coefficients(boxed coefficients) to minimize the error of nonlinear approximation.

In contrast to the spaces V_j the set Σ_n is not a linear space. If one adds two elements from Σ_n the resulting function is in general in Σ_{2n}. It will be necessary to limit this nonlinearity in the following way:

We assume that there exists a constant c such that for all $n \in \mathbb{N}$

$$\Sigma_n + \Sigma_n \subset \Sigma_{cn}. \tag{8.2.3}$$

In comparison to the definition of linear approximation spaces in (5.0.2) we see that the approximation here is of nonlinear type. Thus, we consider the error of nonlinear approximation

$$\operatorname{dist}_X(f, \Sigma_n) := \inf_{S \in \Sigma_n} \|f - S\|_X. \tag{8.2.4}$$

We say then the function S realizes the *best n-term approximation* of f from Σ_n in X. The existence of a best N-term approximation was proved in Dubinin (1997).

To describe a certain decay of the error of nonlinear approximation we define nonlinear approximation spaces as follows.

Definition 8.2.1 (Nonlinear Approximation Spaces). *Let X be a* BANACH *space and $(\Sigma_j)_{j\geq 0}$ a nested sequence of subspaces of X such that their union $\bigcup_{j\geq 0} \Sigma_j$ is dense in X. For real numbers $s > 0, \rho > 1$ and $1 \leq q \leq \infty$ we define the* nonlinear approximation space $\mathcal{NA}^s_q(X)$ *as the set of all functions $f \in X$ such that*

$$|f|_{\mathcal{NA}^s_q(X)} := \begin{cases} \left(\sum_{n=1}^{\infty} (n^s \operatorname{dist}_X(f, \Sigma_n))^q\right)^{1/q} & , 1 \leq q < \infty \\ \sup_{n\geq 1} n^s \operatorname{dist}_X(f, \Sigma_n) & , q = \infty \end{cases} \tag{8.2.5}$$

is finite. The norm is defined as

$$\|f\|_{\mathcal{NA}^s_q(X)} := \|f\|_X + |f|_{\mathcal{NA}^s_q(X)}. \tag{8.2.6}$$

Remark 8.2.1. *Due to the monotonicity of the sequence* $(\operatorname{dist}_X(f, \Sigma_n))_{n\in\mathbb{N}}$ *we obtain an equivalent semi-norm*

$$|f|_{\mathcal{NA}^s_q(X)} \sim \begin{cases} \left(\sum_{j=0}^{\infty} (\rho^{sjd} \operatorname{dist}_X(f, \Sigma_{\rho^{jd}}))^q\right)^{1/q} & , 1 \leq q < \infty \\ \sup_{j\geq 0} \rho^{sjd} \operatorname{dist}_X(f, \Sigma_{\rho^{jd}}) & , q = \infty. \end{cases} \tag{8.2.7}$$

The problem to minimize the error of nonlinear approximation (8.2.4) is easy to solve for nonlinear approximation in $L_2(\mathbb{R}^d)$ and orthogonal wavelet systems. For this purpose we order the coefficients d_λ by their absolute value. We denote by Λ_n the coefficients corresponding to the n largest coefficients. Then the function $S_n := \sum_{\lambda\in\Lambda_n} d_\lambda \psi_\lambda$ minimizes the error (8.2.4). It is not trivial to prove that this strategy also works for approximation in $L_p(\mathbb{R}^d)$ when $p \neq 2$ and for biorthogonal systems. DeVore, Jawerth, and Popov (1992) established the solution to this problem for multivariate dyadic wavelet decompositions. For a description of this let Λ_n denote the coefficients corresponding to the n largest contributions with respect to $\|d_\lambda \psi_\lambda\|_{L_p(\mathbb{R}^d)}$. They showed that under certain conditions on the wavelets ψ one has

$$\|f - S_n\|_{L_p(\mathbb{R}^d)} = \mathcal{O}(n^{-\alpha/d}) \Longleftrightarrow \operatorname{dist}_p(f, \Sigma_n) = \mathcal{O}(n^{-\alpha/d}). \tag{8.2.8}$$

We give one part of this result in Theorem 8.3.2. There we generalize to the more general case of non-separable wavelet decompositions with isotropic scaling matrices.

An interesting question is which functions satisfy the condition (8.2.8). One version of the answer is also presented in the work of DeVore, Jawerth, and Popov (1992). It states that for all τ with $\frac{1}{\tau} = \frac{\alpha}{d} + \frac{1}{p}$ the following estimates are equivalent:

$$\begin{aligned} &(i) \quad \sum_{n=1}^{\infty} (n^{\alpha/d} \operatorname{dist}_p(f, \Sigma_n))^\tau \frac{1}{n} < \infty \\ &(ii) \quad \sum_{n=1}^{\infty} (n^{\alpha/d} \|f - S_n\|_{L_p(\mathbb{R}^d)})^\tau \frac{1}{n} < \infty \\ &(iii) \quad f \in B^\alpha_\tau(L_\tau(\mathbb{R}^d)). \end{aligned} \tag{8.2.9}$$

The main input to establish results as in (8.2.9) is to prove a suitable Jackson and a Bernstein inequality. They are the fundamental components needed to employ the machinery of the real method of interpolation. In the next section we prove such a Jackson and a Bernstein inequality for non-separable wavelet decompositions with isotropic scaling matrices. Then we present a brief introduction into the real method of interpolation in Section 8.4. This technique allows us in a very powerful way to characterize Besov spaces $B^\alpha_\tau(L_\tau(\mathbb{R}^d))$ by approximation spaces with respect to condition (i). Note that these approximation spaces are directly related to our definition in (8.2.5). There are three works we want to emphasize for a brilliant overview about the general framework of nonlinear approximation with wavelet decompositions. More introductory are the surveys of DeVore (1998) and DeVore and Lucier (1992). Many details and complete proofs are collected in Cohen (2003).

We give in Section 8.4 a proof of the equivalence (i)$\Longleftrightarrow$ (ii) in the setting of non-separable wavelet decompositions with isotropic scaling matrices. We follow the line of the proofs of Cohen (2003, Sec. 4.3), where this result is established for biorthogonal multivariate separable wavelet expansions.

8.3 Nonlinear Approximation in L_p Spaces

In this section we prove that a simple thresholding procedure realizes a near *best N-term approximation*. We say a family (B_N), $N \in \mathbb{N}$ of nonlinear operators mapping X to Σ_N realizes a *near best N-term approximation* if there exist a constant $C > 1$ such that

$$\|f - B_N\|_X \leq C \operatorname{dist}_X(f, \Sigma_N) \tag{8.3.1}$$

where C is independent of N.

The first lemma of this section will estimate the amount of contribution of linear combinations of wavelets in dependence of the number of coefficients. The following observation was proved by Temlyakov (1998a) for dyadic wavelet expansions and extended with regard to the multivariate dyadic HAAR system in Temlyakov (1998b). We follow the line of Temlyakov (1998b) to extend the result for non-separable wavelet decompositions.

Lemma 8.3.1. *Let $1 < p < \infty$. Assume that the system of functions $\{\psi_\lambda\}_{\lambda\in\nabla}$ is a wavelet basis generated from a biorthogonal pair of compactly supported continuous scaling functions $(\phi, \tilde{\phi})$ in the sense of (2.1.19). Let $\tilde{\psi}$ be the dual wavelet to ψ. Furthermore, suppose that the supports of ψ_λ do not overlap too much in the following sense:*
There exists a constant $K > 0$ such that

$$\sharp\{\mu \in \nabla_j : \operatorname{supp}\psi_\lambda \cap \operatorname{supp}\psi_\mu \neq \emptyset, \lambda \in \Lambda\} \leq K. \tag{8.3.2}$$

Let $f \in L_p(\mathbb{R}^d)$ have the expansion

$$f = \sum_{\lambda\in\Lambda} d_\lambda \psi_\lambda, \qquad |\Lambda| < \infty. \tag{8.3.3}$$

Let m_1 defined such that for all $\lambda \in \Lambda$

$$\sup_{\lambda\in\Lambda} \|d_\lambda \psi_\lambda\|_{L_p(\mathbb{R}^d)} \leq m_1. \tag{8.3.4}$$

Then it follows

$$\|f\|_{L_p(\mathbb{R}^d)} \lesssim m_1 |\Lambda|^{1/p}. \tag{8.3.5}$$

Similarly, if for all $\lambda \in \Lambda$

$$\inf_{\lambda\in\Lambda} \|d_\lambda \psi_\lambda\|_{L_p(\mathbb{R}^d)} \geq m_2 \tag{8.3.6}$$

then it follows

$$\|f\|_{L_p(\mathbb{R}^d)} \gtrsim m_2 |\Lambda|^{1/p}. \tag{8.3.7}$$

Proof. We start to show (8.3.5). Since $|\Lambda| < \infty$ we get with (2.1.16) that for all $\lambda \in \Lambda$

$$\begin{aligned} \|d_\lambda \psi_\lambda\|_{L_\infty(\mathbb{R}^d)} &\leq |\det M|^{|\lambda|/2} \|d_\lambda \psi\|_{L_\infty(\mathbb{R}^d)} \\ &\lesssim |\det M|^{|\lambda|/2} \|d_\lambda \psi\|_{L_p(\mathbb{R}^d)} \\ &= |\det M|^{|\lambda|/2} |\det M|^{-|\lambda|\frac{1}{2}-\frac{1}{p}} \|d_\lambda \psi_\lambda\|_{L_p(\mathbb{R}^d)} \\ &= |\det M|^{|\lambda|/p} \|d_\lambda \psi_\lambda\|_{L_p(\mathbb{R}^d)} \\ &\lesssim |\det M|^{|\lambda|/p} m_1. \end{aligned}$$

With assumption (8.3.2) we obtain

$$\| \sum_{\lambda \in \Lambda \cap \nabla_j} d_\lambda \psi_\lambda \|_{L_\infty(\mathbb{R}^d)} \lesssim |\det M|^{j/p} m_1. \tag{8.3.8}$$

We define for $x \in \mathbb{R}^d$

$$J(x) := \max\{j \in \mathbb{Z} : \ x \in \operatorname{supp} \psi_\lambda, \lambda \in \Lambda \cap \nabla_j\}. \tag{8.3.9}$$

and the set

$$\Omega_j := \{x \in \mathbb{R}^d : \ J(x) = j\}. \tag{8.3.10}$$

It follows that

$$|\Omega_j| \leq | \bigcup_{\lambda \in \Lambda \cap \nabla_j} \operatorname{supp} \psi_\lambda| \lesssim |\Lambda \cap \nabla_j| \cdot |\operatorname{supp} \psi|^{-j}. \tag{8.3.11}$$

Also, together with equation (8.3.8) we get for $x \in \Omega_j$

$$\begin{aligned} |\sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda(x)| &\leq \sum_{l \leq j} | \sum_{\lambda \in \Lambda \cap \nabla_l} d_\lambda \psi_\lambda(x)| \\ &\lesssim \sum_{l \leq j} |\det M|^{l/p} m_1 \\ &\lesssim |\det M|^{j/p} m_1. \end{aligned}$$

We conclude the proof of (8.3.5) with

$$\begin{aligned} \|\sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda\|^p_{L_p(\mathbb{R}^d)} &= \sum_{j \in \mathbb{Z}} \int_{\Omega_j} |\sum_{\lambda \in \Lambda} d_\lambda \psi_\lambda|^p \\ &\lesssim \sum_{j \in \mathbb{Z}} |\Omega_j| \cdot |\det M|^j (m_1)^p \\ &\lesssim \sum_{j \in \mathbb{Z}} |\Lambda \cap \nabla_j| \cdot |\operatorname{supp} \psi|^{-j} |\det M|^j (m_1)^p \\ &\lesssim |\Lambda| m_1^p. \end{aligned}$$

Now we show inequality (8.3.7). Therefore we define a function

$$g := \sum_{\lambda\in\Lambda} |d_\lambda|^{-2}\bar{d}_\lambda\tilde{\psi}_\lambda. \tag{8.3.12}$$

Then it follows that

$$\begin{aligned}\langle f, g\rangle &= \langle \sum_{\lambda\in\Lambda} d_\lambda\psi_\lambda, \sum_{\lambda\in\Lambda} |d_\lambda|^{-2}\bar{d}_\lambda\tilde{\psi}_\lambda\rangle \\ &= \sum_{\lambda\in\Lambda} d_\lambda|d_\lambda|^{-2}\bar{d}_\lambda = |\Lambda|.\end{aligned}$$

This implies for p, q with $\frac{1}{p}+\frac{1}{q}=1$

$$|\Lambda| = \langle f, g\rangle \le \|f\|_{L_p(\mathbb{R}^d)}\|g\|_{L_q(\mathbb{R}^d)}. \tag{8.3.13}$$

Now we check condition (8.3.4) for the function g. For all $\lambda \in \Lambda$ we see by assumption (8.3.6) and the norm property (2.1.16) that

$$\begin{aligned}\| \, |d_\lambda|^{-2}\bar{d}_\lambda\tilde{\psi}_\lambda\|_{L_q(\mathbb{R}^d)} &\lesssim |d_\lambda|^{-1}|\det M|^{|\lambda|(\frac{1}{2}-\frac{1}{q})}\|\tilde{\psi}\|_{L_q(\mathbb{R}^d)} \\ &\lesssim |d_\lambda|^{-1}|\det M|^{-|\lambda|(\frac{1}{2}-\frac{1}{p})}\|\psi\|^{-1}_{L_p(\mathbb{R}^d)}\|\psi\|_{L_p(\mathbb{R}^d)}\|\tilde{\psi}\|_{L_q(\mathbb{R}^d)} \\ &\lesssim \|d_\lambda\psi_\lambda\|^{-1}_{L_p(\mathbb{R}^d)} < \infty.\end{aligned}$$

Hence, we can apply inequality (8.3.5) to get

$$\|g\|_{L_q(\mathbb{R}^d)} \lesssim \sup_{\lambda\in\Lambda} \| \, |d_\lambda|^{-2}\bar{d}_\lambda\tilde{\psi}_\lambda\|_{L_q(\mathbb{R}^d)}|\Lambda|^{1/q}. \tag{8.3.14}$$

Combination of (8.3.13) and the last inequality yields

$$\|\sum_{\lambda\in\Lambda} d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} \ge |\Lambda|\cdot\|g\|^{-1}_{L_q(\mathbb{R}^d)} \gtrsim m_2|\Lambda|^{1/p}, \tag{8.3.15}$$

where we used as above that $\| \, |d_\lambda|^{-2}\bar{d}_\lambda\tilde{\psi}_\lambda\|^{-1}_{L_q(\mathbb{R}^d)} \gtrsim \|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)}$, which concludes the proof. □

As a consequence of this lemma we prove now a thresholding procedure that realizes a best N-term approximation. It was shown that wavelet thresholding provides such a best N-term approximation for dyadic wavelet expansions in DeVore, Jawerth, and Popov (1992). Newer and simpler techniques for this result was presented in the paper Cohen et al. (2000). We mimic the techniques of their proof to extend this result for non-separable wavelet expansions.

For this we consider a subset Λ_N of Λ with cardinality $|\Lambda_N| = N$. This set shall consist of all indices $\lambda \in \Lambda$ such that $\|d_\lambda \psi_\lambda\|_{L_p(\mathbb{R}^d)} \geq \|d_\mu \psi_\mu\|_{L_p(\mathbb{R}^d)}$ for $\mu \in \Lambda \backslash \Lambda_N$. Thus, Λ_N contains the N largest contributions of the wavelet coefficients in L_p sense.

As a possibility to measure L_p-norms of functions from their wavelet decompositions one often employs the so called *square function* which is defined as

$$S(f) := \left(\sum_{\lambda \in \Lambda} |d_\lambda|^2 |\psi_\lambda|^2 \right). \tag{8.3.16}$$

To do this it can be shown from general results in LITTLEWOOD-PALEY theory for $1 < p < \infty$

$$\|S(f)\|_{L_p(\mathbb{R}^d)} \sim \|f\|_{L_p(\mathbb{R}^d)}, \tag{8.3.17}$$

see Meyer (1992). If $p \leq 1$ a similar result holds for the HARDY spaces H_p instead of L_p, see Cohen et al. (2000).

Theorem 8.3.2 (Thresholding). *Assume that for $1 < p < \infty$ the system of functions $\{\psi_\lambda\}_{\lambda \in \nabla}$ is a wavelet basis generated from a biorthogonal pair of compactly supported continuous scaling functions $(\phi, \tilde{\phi})$ in the sense of (2.1.19). Let $\tilde{\psi}$ be the dual wavelet to ψ. Let $\psi \in C^s$ for a small $s > 0$. Furthermore, suppose that the ψ_λ satisfy the overlapping condition (8.3.2). If $f \in L_p(\mathbb{R}^d)$ has the expansion*

$$f = \sum_{\lambda \in \nabla} d_\lambda \psi_\lambda, \tag{8.3.18}$$

then

$$\|f - \sum_{\lambda \in \Lambda_N} d_\lambda \psi_\lambda\|_{L_p(\mathbb{R}^d)} \lesssim \operatorname{dist}_p(f, \Sigma_N). \tag{8.3.19}$$

Proof. Let $B_N f := \sum_{\lambda \in \Lambda_b} b_\lambda \psi_\lambda$ with $|\Lambda_b| \leq N$ and its expansion coefficients b_λ be a best N-term approximation of f in $L_p(\mathbb{R}^d)$ from Σ_N. The existence of a best N-term approximation was proved in Dubinin (1997). We denote $\tilde{B}_N f$ the expansion of $B_N f$ with b_λ replaced by d_λ which also belongs to Σ_N. By definition of Λ_N (coefficients are largest L_p contribution) it follows $S(f - \tilde{B}_N f) \leq S(f - B_N f)$ for the square function defined in (8.3.16). Since $B_N f$ is a best N-term approximation we get together with the equivalence (8.3.17)

$$\|f - \tilde{B}_N f)\|_{L_p(\mathbb{R}^d)} \lesssim \|f - B_N f\|_{L_p(\mathbb{R}^d)} \lesssim \operatorname{dist}_p(f, \Sigma_N). \tag{8.3.20}$$

We define for the function f the associated projection operator T_Λ by

$$T_\Lambda f := \sum_{\lambda\in\Lambda} \langle f, \tilde{\psi}_\lambda\rangle \psi_\lambda. \tag{8.3.21}$$

We find out that

$$\|f - T_{\Lambda_N} f\|_{L_p(\mathbb{R}^d)} \leq \|f - \tilde{B}_N f\|_{L_p(\mathbb{R}^d)} + \|\tilde{B}_N f - T_{\Lambda_N} f\|_{L_p(\mathbb{R}^d)}. \tag{8.3.22}$$

It is therefore enough to bound

$$\begin{aligned} \tilde{B}_N f - T_{\Lambda_N} f &= \sum_{\lambda\in\Lambda_b} d_\lambda \psi_\lambda - \sum_{\lambda\in\Lambda_N} d_\lambda\psi_\lambda && (8.3.23)\\ &= \sum_{\lambda\in\Lambda_b\setminus\Lambda_N} d_\lambda\psi_\lambda - \sum_{\lambda\in\Lambda_N\setminus\Lambda_b} d_\lambda\psi_\lambda && (8.3.24)\\ &=: f_0 + f_1. && (8.3.25) \end{aligned}$$

With inequality (8.3.20) follows

$$\begin{aligned} \|f_1\|_{L_p(\mathbb{R}^d)} &= \|\sum_{\lambda\in\Lambda_N\setminus\Lambda_b} d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} && (8.3.26)\\ &\leq \|\sum_{\lambda\in\Lambda\setminus\Lambda_b} d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} && (8.3.27)\\ &= \|\sum_{\lambda\in\Lambda} d_\lambda\psi_\lambda - \sum_{\lambda\in\Lambda_b} d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} && (8.3.28)\\ &= \|f - \tilde{B}_N f\|_{L_p(\mathbb{R}^d)} && (8.3.29)\\ &\lesssim \operatorname{dist}_p(f, \Sigma_N). && (8.3.30) \end{aligned}$$

To bound $\|f_0\|_{L_p(\mathbb{R}^d)}$ we let $m := \inf_{\lambda\in\Lambda_N} \|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)}$. It follows therefore that $\inf_{\lambda\in\Lambda_b\setminus\Lambda_N} \|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} \leq m$. We apply Lemma 8.3.1 to get

$$\|f_0\|_{L_p(\mathbb{R}^d)} \lesssim m|\Lambda_b \setminus \Lambda_N|^{1/p}. \tag{8.3.31}$$

On the other hand we obtain for all $\lambda \in \Lambda_N \setminus \Lambda_b$ that

$$\|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} \geq m$$

and hence again with Lemma 8.3.1 it follows

$$\|f_1\|_{L_p(\mathbb{R}^d)} \gtrsim m|\Lambda_N \setminus \Lambda_b|^{1/p}. \tag{8.3.32}$$

We observe that $|\Lambda_N \setminus \Lambda_b| = |\Lambda_b \setminus \Lambda_N|$. Finally we get

$$\|f_0\|_{L_p(\mathbb{R}^d)} \lesssim \|f_1\|_{L_p(\mathbb{R}^d)} \lesssim \operatorname{dist}_p(f, \Sigma_N)$$

which concludes the proof. $\square$

The next theorem presents a suitable JACKSON inequality for N-term approximation with non-separable wavelet decompositions with isotropic scaling matrices. We need this type of JACKSON inequality for the characterization of nonlinear approximation spaces in the next section.

Theorem 8.3.3 (Jackson Inequality). *Assume that for* $1 < p < \infty$ $\{\psi_\lambda\}_{\lambda\in\nabla}$ *is a wavelet basis generated from a biorthogonal pair of compactly supported continuous scaling functions* $(\phi, \tilde{\phi})$ *in the sense of (2.1.19). Let* $\tilde{\psi}$ *be the dual wavelet to* ψ. *Assume that for any function* $f = \sum_{\lambda\in\nabla} d_\lambda\psi_\lambda$ *in* $B^s_\tau(L_\tau(\mathbb{R}^d))$ *such that* $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$ *and* $s > 0$ *the characterization (5.2.3) holds. Then one has the* JACKSON *type inequality*

$$\operatorname{dist}_p(f, \Sigma_N) \lesssim N^{-s/d}\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))}, \qquad N \in \mathbb{N}. \tag{8.3.33}$$

Proof. Let us denote by $\{\psi_\lambda\}_{\lambda\in\nabla_{j_0-1}}$ the scaling function basis $\{\phi_\lambda\}_{\lambda\in\mathbb{Z}^d}$ on the coarsest level j_0 with a similar convention for the dual scaling functions. If we assume that $j_0 = 0$ (since all other cases follow by rescaling) we can rewrite the norm equivalence (5.2.3) as

$$\begin{aligned} \|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))} &\sim \Big(\sum_{j\geq -1} (\rho^{js}\|Q_j f\|_{L_\tau(\mathbb{R}^d)})^\tau\Big)^{1/\tau} & (8.3.34)\\ &\sim \Big(\sum_{j\geq -1} (\rho^{js}|\det M|^{j(\frac{1}{2}-\frac{1}{\tau})}\|d^{(j)}\|_{\ell_\tau})^\tau\Big)^{1/\tau}. & (8.3.35) \end{aligned}$$

Using that $s = d(\frac{1}{\tau} - \frac{1}{p})$ and the fact that $\rho^d = |\det M|$ we obtain with the rescaling argument (2.1.16) and the stability property (2.4.7) the equivalence

$$\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))} \sim \Big(\sum_{\lambda\in\nabla} \|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)})^\tau\Big)^{1/\tau}. \tag{8.3.36}$$

Note that for $\tau \geq 1$ this result is just a rephrasing of the norm equivalence (5.2.3). If $\tau < 1$ the coefficients are not obviously related to the BESOV norm. Using the characterization $|f|_{B^s_\tau(L_\tau(\mathbb{R}^d))} \sim |f|_{\mathcal{A}^s_\tau(L_p(\mathbb{R}^d))}$ shows that this holds for the whole range of τ, see Canuto and Tabacco (1997, Sec. 6.6).

We introduce the space *weak*-ℓ_τ

$$\ell^\tau_w := \{(a_\lambda)_{\lambda\in\mathbb{Z}^d} : \ \sharp\{\lambda : |a_\lambda| \geq \epsilon\} \leq \epsilon^{-q} \text{ for all } \epsilon > 0\}. \tag{8.3.37}$$

We show that the sequence $(\|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)})_{\lambda\in\nabla}$ is in weak-ℓ_τ. For all $\epsilon > 0$ it follows with equivalence (8.3.36)

$$\|f\|^\tau_{B^s_\tau(L_\tau(\mathbb{R}^d))} \gtrsim \sharp\{\lambda : \|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} \geq \epsilon\}\epsilon^\tau. \tag{8.3.38}$$

We consider the set

$$A_j := \{\lambda : \, C\rho^{-jd/\tau}\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))} \le \|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)} \le C\rho^{-(j-1)d/\tau}\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))}\}. \tag{8.3.39}$$

With (8.3.38) we see that there exists such a constant C depending on the one in (8.3.36) such that $|A_j| \le \rho^{jd}$. We use again Lemma 8.3.1, in particular inequality (8.3.5) to get for T_{A_j} defined in (8.3.21)

$$\begin{aligned} \|T_{A_j} f\|_{L_p(\mathbb{R}^d)} &\lesssim \sup_{\lambda\in A_j} \|d_\lambda\psi_\lambda\|_{L_p(\mathbb{R}^d)}|A_j|^{1/p} \\ &\lesssim \rho^{-(j-1)d/\tau}\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))}\rho^{jd/p} \\ &\lesssim \rho^{-js}\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))}. \end{aligned}$$

We define $B_j := \bigcup_{l=0}^{j-1}$ and derive $|B_j| \le \rho^{jd}$ since $|A_l| \le \rho^{ld}$. For $N = \rho^{jd}$ we conclude the proof of the JACKSON type estimate with the remark that $\mathrm{dist}_p(f, \Sigma_N)$ is a monotone function in N by estimation of

$$\begin{aligned} \mathrm{dist}_p(f, \Sigma_N) &\lesssim \|f - T_{B_j} f\|_{L_p(\mathbb{R}^d)} \\ &\lesssim \sum_{l\ge j} \|T_{A_l} f\|_{L_p(\mathbb{R}^d)} \\ &\lesssim \sum_{l\ge j} \rho^{-ls}\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))} \\ &\lesssim N^{-s/d}\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))}. \end{aligned}$$

□

The second important component that we need for the characterization of nonlinear approximation spaces is a suitable BERNSTEIN inequality which is proved in the next theorem.

Theorem 8.3.4 (Bernstein Inequality). *Assume that for* $1 < p < \infty$ $\{\psi_\lambda\}_{\lambda\in\nabla}$ *is a wavelet basis generated from a biorthogonal pair of compactly supported continuous scaling functions* $(\phi, \tilde{\phi})$ *in the sense of (2.1.19). Let* $\tilde{\psi}$ *be the dual wavelet to* ψ. *Assume that the function* $f \in \Sigma_N$, *that is* $f = \sum_{\lambda\in\Lambda} d_\lambda\psi_\lambda$ *with* $|\Lambda| \le N$. *Further, let* $f \in B^s_\tau(L_\tau(\mathbb{R}^d))$ *such that* $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$ *and* $s > 0$ *the characterization (5.2.3) holds. Then one has the* BERNSTEIN *type inequality*

$$\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))} \lesssim N^{s/d}\|f\|_{L_p(\mathbb{R}^d)} \tag{8.3.40}$$

Proof. For this proof we employ again the square function. In order to prove the BERNSTEIN inequality we rewrite the norm equivalence (5.2.3) in the same way as in the proof of Theorem 8.3.3 and use the rescaling property of the norm of ψ_λ to get

$$\begin{aligned}
\|f\|^\tau_{B^s_\tau(L_\tau(\mathbb{R}^d))} &\lesssim \sum_{\lambda\in\Lambda} \|d_\lambda \psi_\lambda\|^\tau_{L_p(\mathbb{R}^d)} \\
&= \sum_{\lambda\in\Lambda} |d_\lambda|^\tau \|\psi_\lambda\|^p_{L_p(\mathbb{R}^d)} \|\psi_\lambda\|^{\tau-p}_{L_p(\mathbb{R}^d)} \\
&= \int_{\mathbb{R}^d} \sum_{\lambda\in\Lambda} |d_\lambda|^\tau |\psi_\lambda(x)|^p |\det M|^{|\lambda|(\frac{1}{2}-\frac{1}{p})(\tau-p)} \|\psi\|^{\tau-p}_{L_p(\mathbb{R}^d)} \mathrm{d}x \\
&\lesssim \int_{\mathbb{R}^d} \sum_{\lambda\in\Lambda} |d_\lambda|^\tau |\psi_\lambda(x)|^\tau (|\det M|^{|\lambda|(\frac{1}{p}-\frac{1}{2})(\tau-p)} |\psi_\lambda(x)|)^{p-\tau} \mathrm{d}x \\
&\lesssim \int_{\mathbb{R}^d} (Sf(x))^\tau R_\Lambda(x) \mathrm{d}x,
\end{aligned}$$

where

$$R_\Lambda(x) = \left(\sum_{\lambda\in\Lambda} (|\det M|^{|\lambda|(\frac{1}{p}-\frac{1}{2})} |\psi_\lambda(x)|)^{\frac{2(p-\tau)}{2-\tau}} \right)^{\frac{2-\tau}{2}}. \tag{8.3.41}$$

In the last inequality we applied a discrete HÖLDER inequality for $\frac{\tau}{2} + \frac{2-\tau}{2} = 1$.

We recall the definition of $J(x)$ in (8.3.9) and continue by estimating $R_\Lambda(x)$, that is

$$\begin{aligned}
R_\Lambda(x) &\lesssim \left(\sum_{\substack{\lambda\in\Lambda \\ \psi_\lambda(x)\neq 0}} (|\det M|^{|\lambda|(\frac{1}{p}-\frac{1}{2})} |\det M|^{|\lambda|/2})^{\frac{2(p-\tau)}{2-\tau}} \right)^{\frac{2-\tau}{2}} \\
&\lesssim \left(\sum_{\substack{\lambda\in\Lambda \\ \psi_\lambda(x)\neq 0}} |\det M|^{2|\lambda|\frac{1}{p}\frac{(p-\tau)}{2-\tau}} \right)^{\frac{2-\tau}{2}} \\
&\lesssim |\det M|^{J(x)(1-\frac{\tau}{p})}.
\end{aligned}$$

Now we employ again Ω_j as defined in (8.3.10). We use Hölder inequality for functions with $\frac{\tau}{p} + \frac{p-\tau}{p} = 1$ and the equivalence (8.3.17) to deduce

$$\begin{aligned}
\|f\|^{\tau}_{B^s_\tau(L_\tau(\mathbb{R}^d))} &\lesssim \|Sf\|^{\tau/p}_{L_p(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} |\det M|^{J(x)(1-\frac{\tau}{p})\frac{p}{p-\tau}} \right)^{\frac{p-\tau}{p}} \\
&\lesssim \|f\|^{\tau}_{L_p(\mathbb{R}^d)} \left(\sum_{j\geq -1} |\Omega_j| \cdot |\det M|^j \right)^{\frac{p-\tau}{p}} \\
&\lesssim \|f\|^{\tau}_{L_p(\mathbb{R}^d)} \left(\sum_{j\geq -1} |\Lambda \cap \nabla_j| \cdot |\operatorname{supp}\psi|^{-j} |\det M|^j \right)^{\frac{p-\tau}{p}} \\
&\lesssim \|f\|^{\tau}_{L_p(\mathbb{R}^d)} N^{\frac{p-\tau}{p}}.
\end{aligned}$$

We conclude the proof with

$$\|f\|_{B^s_\tau(L_\tau(\mathbb{R}^d))} \lesssim \|f\|_{L_p(\mathbb{R}^d)} N^{\frac{1}{\tau}-\frac{1}{p}} = \|f\|_{L_p(\mathbb{R}^d)} N^{\frac{s}{d}}.$$

□

8.4 Nonlinear Approximation Spaces and Interpolation Spaces

Now we involve a more general mechanism to describe the nonlinear approximation spaces. We make use of the *real method of interpolation theory*. Interpolation spaces are defined via K-functionals. To explain this we let X and Y be quasi-normed linear spaces. We assume that Y is continuously embedded in X, i.e. $X \subset Y$ with $\|\cdot\|_X \lesssim \|\cdot\|_Y$. We define for any $t > 0$ the K-functional

$$K(f,t) := K(f,t,X,Y) := \inf_{g\in Y} \|f-g\|_X + t|g|_Y \tag{8.4.1}$$

with a norm $\|\cdot\|_X$ on X and a quasi-semi-norm $|\cdot\|_Y$ on Y.

K-functionals have many uses. For instance they are used to describe a certain type of approximation. That is to find a function g that minimizes the distance in X with a penalty term $|g|_Y$.

Originally they were introduced as terms to generate *interpolation spaces*. This is also the reason why we consider them here. The most common definition of interpolation spaces is given by θ, q norms.

Definition 8.4.1 (Interpolation spaces). *The interpolation space* $(X,Y)_{\theta,q}$, $\theta > 0, 0 < q \le \infty$ *is the set of all functions* $f \in X$ *such that*

$$|f|_{(X,Y)_{\theta,q}} := \begin{cases} \left(\int_0^\infty [t^{-\theta}K(f,t)]^q \frac{\mathrm{d}t}{t}\right)^{\frac{1}{q}} & \text{when } 1 \le q < \infty, \\ \sup_{0<t<\infty} t^{-\theta}K(f,t) & q = \infty \end{cases} \tag{8.4.2}$$

is finite.

Remark 8.4.1. *It is possible to build interpolation spaces between a pair interpolation spaces. The resulting spaces remain interpolation spaces of the original ones. The precise formulation of this fact is described by the famous* Reiteration Theorem of interpolation *(see e.g. Bergh and Löfström (1976)). It states the following:*

Let $X_1 := (X,Y)_{\theta_1,q_1}$ *and* $X_2 := (X,Y)_{\theta_2,q_2}$. *Then one has for all* $0 < \theta < 1$ *and* $0 < q < \infty$

$$(X_1, X_2)_{\theta,q} = (X,Y)_{\alpha,q} \tag{8.4.3}$$

with $\alpha := (1-\theta)\theta_1 + \theta\theta_2$.

Remark 8.4.2. *In DeVore and Lorentz (1993, Chap.6,§7) it is shown that the continuous norm in (8.4.2) can be discretized similar to the discrete* Besov *semi-norm in Proposition 2.5.1. Namely, for a function* $f \in X$ *and a fixed positive number* $\rho > 1$ *one has*

$$|f|_{(X,Y)_{\theta,q}} \sim \begin{cases} \left(\sum_{j=0}^\infty (\rho^{j\theta}K(f,\rho^{-j}))^q\right)^{\frac{1}{q}} & \text{when } 1 \le q < \infty, \\ \sup_{0<t<\infty} \rho^{j\theta}K(f,\rho^{-j}) & q = \infty. \end{cases} \tag{8.4.4}$$

We see that in this form of semi-norm the definition of interpolation spaces and approximation spaces are almost identical. If we prove that

$$\inf_{g \in \Sigma_{\rho^{jd}}} \|f - g\|_{L_p(\mathbb{R}^d)} \sim K(f, \rho^{-j}, X, Y), \tag{8.4.5}$$

then these two spaces coincide. That is, one space can be characterized by the other. To show the above equivalence (8.4.5) we have to make the „right choice“of the spaces X and Y. It turns out that this choice is determined by the Jackson and Bernstein inequalities that were previously shown. In DeVore and Lorentz (1993, Chap.7) it is shown that if the Jackson inequality

$$\mathrm{dist}_X(f, \Sigma_N) \lesssim N^{-r/d}|f|_Y, \qquad N = 1, 2, \ldots \tag{8.4.6}$$

and the BERNSTEIN inequality

$$\|f\|_Y \lesssim N^{r/d}\|f\|X, \qquad f \in \Sigma_N, N = 1, 2, \ldots \tag{8.4.7}$$

hold, then one has for $0 < s < r$ and $0 < q < \infty$

$$\mathcal{N}\mathcal{A}_q^{s/d}(X) = (X, Y)_{s/r,q}. \tag{8.4.8}$$

In particular, for the specific versions of the JACKSON inequality (8.3.33) and the BERNSTEIN inequality (8.3.40) with $X = L_p(\mathbb{R}^d)$ and $Y = B_\tau^s(L_\tau(\mathbb{R}^d))$, $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$ we obtain for any $0 < \gamma < s$ and $0 < q \leq \infty$,

$$\mathcal{N}\mathcal{A}_q^{\gamma/d}(L_p(\mathbb{R}^d)) = (L_p(\mathbb{R}^d), B_\tau^s(L_\tau(\mathbb{R}^d)))_{\gamma/s,q}. \tag{8.4.9}$$

In DeVore (1998, Sec. 7.4, Remark(ii)) it is remarked that for each γ there is one value of q where the interpolation space on the right hand side of (8.4.9) is a BESOV space. The full proof is presented in DeVore and Popov (1988). This is valid for $\frac{1}{q} = \frac{\gamma}{d} + \frac{1}{p}$ such that we finally get the equivalence

$$\mathcal{N}\mathcal{A}_q^{\gamma/d}(L_p(\mathbb{R}^d)) = B_q^\gamma(L_q(\mathbb{R}^d)) \tag{8.4.10}$$

with equivalent norms.

Remark 8.4.3. *To prove the characterization (8.4.10) we first established a* JACKSON *and a* BERNSTEIN *inequality. Then we used the real method of interpolation. In Chapter 5,6 and 7 we could circumvent to apply this method by directly proving the norm equivalence between discrete* BESOV *norms and weighted sequences of approximation errors or wavelet coefficients. Intrinsically we used for the proofs of the norm equivalences in Chapter 5,6 and 7 the same technique as for the proof of the more general interpolation result (8.4.9). This can be achieved by direct comparison of the K-functional $K(f, \rho^{-j})$ and the modulus of smoothness $\omega_l(f, \rho^{-j})_p$ (resp. the error of approximation $\mathrm{dist}_p(f, V_j)$) as we did in e.g. in Theorem 5.2.1 or Theorem 7.2.1. In this chapter we had to have compared the terms $K(f, \rho^{-j}, L_p(\mathbb{R}^d), B_\tau^s(L_\tau(\mathbb{R}^d)))$ and $\omega_l(f, \rho^{-j})_\tau$ (resp. the error of nonlinear approximation $\mathrm{dist}_p(f, \Sigma_j)$) for $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$. This turns out to be much more challenging due to the different norms in p and τ. To the authors knowledge there is no reference where these terms are directly compared within this setting.*

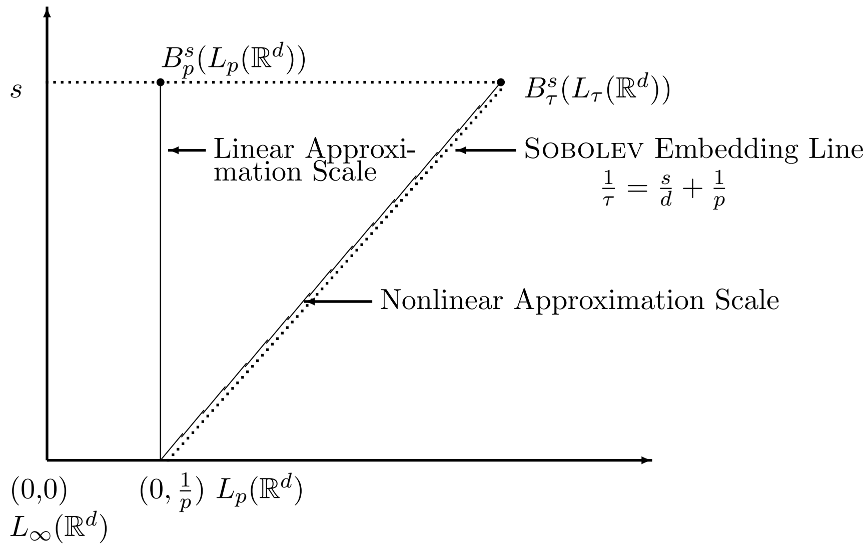


Figure 8.2: Graphical interpretation of the difference of linear and nonlinear approximation scale together with the SOBOLEV embedding line.

Remark 8.4.4. *In Theorem 5.2.1 we have seen that functions with smoothness of order α measured in $L_p(\mathbb{R}^d)$ can be characterized by exhibiting approximation rate $\mathcal{O}(\rho^{-js})$ in $L_p(\mathbb{R}^d)$ for a fixed p. The characterization (8.4.10) shows that we obtain for nonlinear approximation with a given smoothness order α measured in $L_\tau(\mathbb{R}^d)$ the same approximation rate in $L_p(\mathbb{R}^d)$ where $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$, see Figure 8.2 for a graphical visualization of this result. . Since $\tau < p$ we have a weaker smoothness conditions on the underlying function when we approximate with nonlinear methods. Moreover, if the smoothness index s becomes large τ might be less than one. It is known that as τ approaches zero the ℓ_τ norm measures the* sparsity *of its argument. That is we could expect that the $B^s_\tau(L_\tau(\mathbb{R}^d))$ norm is concentrated on a small number of (wavelet)coefficients. This is exactly the situation if the underlying function has some isolated singularities and smooth parts otherwise. Then this function has a* sparse *wavelet decomposition. Thus we can gain more smoothness by decreasing the norm in which we measure it and consequently achieve a higher rate of approximation. To summarize, functions with singularities can be handled optimal by nonlinear methods.*

Summary

This thesis is a contribution to the relation between BESOV spaces and weighted sequences of wavelet coefficients. The aim is to extend the characterization of BESOV spaces via separable wavelet decompositions to the characterization by non-separable wavelet bases with isotropic scaling matrices. The main focus is set on establishing norm equivalences between discrete BESOV norms and weighted sequences of discrete wavelet coefficients. Norm equivalences have been already established for separable wavelet bases, see e.g. DeVore, Jawerth, and Popov (1992) or an extensive study with detailed proofs and applications to numerical analysis in Cohen (2003).

First we showed in Section 2.4 that under very mild conditions the basis $\Phi_j := \{\phi_{j,k}, k \in \mathbb{Z}^d\}$ and $\Psi_j := \{\psi_{e,j,k}, e \in E, k \in \mathbb{Z}^d\}$ constitute a stable basis. Also we derived that for $1 \leq p \leq \infty$ the projectors P_j and Q_j defined as

$$P_j f = \sum_{k\in\mathbb{Z}^d} \langle f, \tilde{\phi}_{j,k}\rangle \phi_{j,k}$$

and

$$Q_j f = \sum_{e\in E} \sum_{k\in\mathbb{Z}^d} \langle f, \tilde{\psi}_{e,j,k}\rangle \psi_{e,j,k}$$

are uniformly bounded in $L_p(\mathbb{R}^d)$.

In Section 3.3 we derived a JACKSON type estimate. The full relation between smoothness of a function and the corresponding approximation rate by shift-invariant spaces is written in the following theorem.

Theorem (Jackson Type Estimate). *Let $f \in W^{r+1}(L_p(\mathbb{R}^d))$. Further, let $\phi \in C_0(\mathbb{R}^d)$, $\phi \in W^r(L_1(\mathbb{R}^d))$ be a $(\mathbf{a}, M)$-refinable ℓ_2-stable function. If ϕ satisfies the generalized* STRANG-FIX *conditions of order r, i.e.*

$$(\partial^\alpha \hat{\phi})(2\pi k) = 0, \quad \textit{for all}\, k \in \mathbb{Z}^d \setminus \{0\},\ |\alpha| \leq r,$$

then we have the JACKSON *type inequality*

$$\inf_{g \in V_j} \|f - g\|_{L_p(\mathbb{R}^d)} \lesssim \|M^{-j}\|_\infty^{r+1} |f|_{W^{r+1}(L_p(\mathbb{R}^d))}.$$

In Section 4.1 we affiliated a sharp BERNSTEIN estimate that bound the smoothness norm of functions belonging to the spaces of the multiresolution analysis.

Theorem (General Bernstein inequality). *Let* $1 \leq p \leq \infty$*. Assume that* $\phi \in L_p(\mathbb{R}^d)$ *and* $\tilde{\phi} \in L_{p'}(\mathbb{R}^d)$*, where* $1 = \frac{1}{p} + \frac{1}{p'}$*, are* $(\mathbf{a}, M)$*-refinable functions. Further, let* $\phi, \tilde{\phi} \in W^{k+1}(\mathbb{R}^d)$*. Then there exists a real number* $\delta > 0$ *such that for each* $g \in V_m$

$$\omega_{k+1}(g; t)_p \lesssim [\min\{1, t\rho^m\}]^{k+\delta} \|g\|_{L_p(\mathbb{R}^d)}.$$

We proved that norm equivalences also hold for non-separable wavelet bases with isotropic scaling matrices. In particular we have shown in Chapter 5, Section 5.2 the following theorem.

Theorem (Norm Equivalences). *Let* $\phi \in C_0^k(\mathbb{R}^d)$ *be a* $(\mathbf{h}, M)$*-refinable function associated with a multiresolution analysis with an isotropic scaling matrix* M *and* ρ *denoting the modulus of its eigenvalues and* $\mathbf{h} \in \ell_1(\mathbb{Z}^d)$*. Furthermore, let* ϕ *be a* ℓ_2*-stable generator of a multiresolution analysis of* $L_2(\mathbb{R}^d)$ *and the assumptions of the* JACKSON *inequality and the* BERNSTEIN *inequality hold.*

Then, for $0 < \alpha < k + \delta$*,* $\delta > 0$ *small enough, the following conditions are equivalent:*

$$(i) \qquad \left(\sum_{j=-1}^{\infty} (\rho^{j\alpha} \|f - P_j f\|_{L_p(\mathbb{R}^d)})^q \right)^{\frac{1}{q}} < \infty$$

$$(ii) \qquad \left(\sum_{j=-1}^{\infty} (\rho^{j\alpha} \|Q_j f\|_{L_p(\mathbb{R}^d)})^q \right)^{\frac{1}{q}} < \infty$$

$$(iii) \qquad \left(\sum_{k \in \mathbb{Z}^d} |\langle f, \tilde{\phi}(\cdot - k) \rangle|^p \right)^{\frac{1}{p}} + \left(\sum_{j=0}^{\infty} (\rho^{j\alpha} |\det M|^{j(\frac{1}{2} - \frac{1}{p})} \|d^{(j)}\|_{\ell_p})^q \right)^{\frac{1}{q}} < \infty$$

$$(iv) \qquad f \in \mathcal{A}_q^\alpha(L_p(\mathbb{R}^d))$$

$$(v) \qquad \|f\|_{(\alpha, \infty, p, q)} < \infty$$

$$(vi) \qquad f \in B_q^\alpha(L_p(\mathbb{R}^d)).$$

We showed that these equivalences also hold for negative smoothness (cf. Chapter 6).

For the field of nonlinear approximation which was considered in Chapter 8 we also needed to extend the range of consideration to values $p < 1$ of the norm in which we measure the smoothness. This is also known as *unstable approximation*. We presented a result that provides a characterization also for this range, see Chapter 7.

With the norm equivalences in hand we considered N-term approximation in $L_p(\mathbb{R}^d)$ for non-separable wavelet bases with isotropic scaling matrices. We showed in Chapter 8 that the corresponding nonlinear approximation spaces can be characterized by certain Besov spaces. To be precise, we get for $\frac{1}{\tau} = \frac{s}{d} + \frac{1}{p}$ and for any $0 < \gamma < s$ and $0 < q \leq \infty$,

$$\mathcal{N}\mathcal{A}_q^{\gamma/d}(L_p(\mathbb{R}^d)) = (L_p(\mathbb{R}^d), B_\tau^s(L_\tau(\mathbb{R}^d)))_{\gamma/s,q}.$$

It is known from DeVore (1998) that for each γ there is one value of q where the interpolation space on the right hand side is a Besov space. We concluded that for non-separable wavelet bases with isotropic scaling matrices we have

$$\mathcal{N}\mathcal{A}_q^{\gamma/d}(L_p(\mathbb{R}^d)) = B_q^\gamma(L_q(\mathbb{R}^d))$$

with equivalent norms which are essentially the same scale of Besov spaces as for separable wavelet decompositions.

We have thus extended the well known results about the characterization of Besov spaces via dyadic wavelet expansions are extended for those cases where the dilation is given by a general expanding isotropic integer matrix. It turns out that we can characterize the same scale of Besov spaces for linear as well as for nonlinear approximation with both separable and non-separable wavelet decompositions with isotropic scaling matrices.

Bibliography

Adams, R.A. 1978. *Sobolev Spaces.* Boston, San Diego, New York: Academic press.

Andrews, R., and D. Nguyen. 1998. "Separable and Quincunx Wavelet Image Coding." *6 th IEEE Intern. Workshop on Signal Processing and Communication Systems, Melbourne, Australia.*

Andrews, R., and D.T. Nguyen. 2002. "Separable Versus Quincunx Wavelet Transforms For Image Compression." Technical Report, Department of Electrical Engeneering and Computer Science, University of Tasmania.

Barinka, A., T. Barsch, P. Charton, A. Cohen, S. Dahlke, W. Dahmen, and K. Urban. 2001. "Adaptive wavelet schemes for elliptic problems–implementation and numerical experiments." *SIAM J. Sci. Comput.* 23 (3): 910–939.

Belogay, E., and Y. Wang. 1999. "Arbitrarily smooth orthogonal nonseparable wavelets in $\mathbb{R}^2$." *SIAM J. Math. Anal.* 30 (3): 678–697.

Bergh, J., and J. Löfström. 1976. *Interpolation spaces. An introduction.* Berlin: Springer-Verlag. Grundlehren der Mathematischen Wissenschaften, No. 223.

Bonnet, S., F. Peyrin, F. Turjman, and R. Prost. 2000. "Tomographic reconstruction using nonseparable wavelets." *IEEE Trans. Image Process.* 9 (8): 1445–1450.

Bownik, M. 2001. "The construction of r-regular wavelets for arbitrary dilations." *J. Fourier Anal. Appl.* 7:489–506.

Brudnyi, Y. 1970. "Approximation of functions of n–variables by quasi polynomials." *Math. USSR Izv. 4*, pp. 568–586.

Canuto, C., and A. Tabacco. 1997. “Multilevel decompositions of functional spaces.” *J. Fourier Anal. Appl.* 3 (6): 715–742.

Cavaretta, A.S., W. Dahmen, and C.A. Micchelli. 1991. “Stationary Subdivision.” *Mem. Amer. Math. Soc.* 93:1–186.

Cheung, H.L., C. Tang, and D.-X. Zhou. 2002. “Supports of locally linearly independent M-refinable functions, attractors of iterated function systems and tilings.” *Adv. Comput. Math.* 17 (3): 257–268.

Ciarlet, P.G. 1991. Pages 17–351 in *Basic error estimate for the finite element method*, Volume 2. Elsevier, Amsterdam.

Cohen, A. 2003. *Numerical analysis of wavelet methods.* Studies in Mathematics and its Applications 32. Amsterdam: North-Holland. xviii.

Cohen, A., and I. Daubechies. 1993. “Non-separable bidimensional wavelet bases.” *Rev. Mat. Iberoam.* 9 (1): 51–137.

Cohen, A., I. Daubechies, and J.-C. Feauveau. 1992. “Biorthogonal bases of compactly supported wavelets.” *Comm. Pure Appl. Math.* 45 (5): 485–560.

Cohen, A., DeVore, R.A., and R. Hochmuth. 2000. “Restricted nonlinear approximation.” *Constr. Approx.* 16 (1): 85–113.

Cohen, A., K. Gröchenig, and Lars F. Villemoes. 1999. “Regularity of Multivariate Refinable Functions.” *Constructive Approximation* 15 (2): 241–255.

Dahlke, S. 1997. *Wavelets: Construction Principles and Application to the Numerical Treatment of Operator Equations.* Aachen: Shaker-Verlag.

———. 1999a. “Besov regularity for elliptic boundary value problems in polygonal domains.” *Appl. Math. Lett.* 12 (6): 31–36.

———. 1999b. “Besov regularity for the Stokes problem.” In *Advances in multivariate approximation. Proceedings of the 3rd international conference on multivariate approximation theory, Witten-Bommerholz, Germany, September 27-October 2, 1998.*, edited by Werner et al. Haumann, 129–138. Wiley-VCH, Berlin.

———. 2003. “Besov regularity for the Neumann problem.” In *Function spaces, differential operators and nonlinear analysis. The Hans Triebel anniversary volume. Based on the lectures given at the international conference on function spaces, differential operators and nonlinear analysis, FSDONA-01, Teistungen, Germany, June 28-July 4, 2001, in* *honor of the 65th birthday of H. J. Triebel*, edited by D. Haroske et al., 267–277. Birkhuser, Basel.

Dahlke, S., W. Dahmen, and R.A. DeVore. 1997. “Nonlinear Approximation and Adaptive Techniques for Solving Elliptic Operator Equations.” In *Multicale Wavelet Methods for Partial Differential Equations*, edited by W. Dahmen, A. Kurdila, and P. Oswald, 237–283. Academic Press, San Diego.

Dahlke, S., W. Dahmen, and K. Urban. 2002. “Adaptive wavelet methods for saddle point problems – optimal convergence rates.” *SIAM J. Numer. Anal.* 40 (4): 1230–1262.

Dahlke, S., and R.A. DeVore. 1997. “Besov regularity for elliptic boundary value problems.” *Commun. Partial Diff. Eqns.* 22:1–16.

Dahlke, S., R. Hochmuth, and K. Urban. 2000. “Convergent adaptive wavelet methods for the Stokes problem.” In *Multigrid methods, VI (Gent, 1999)*, Volume 14 of *Lect. Notes Comput. Sci. Eng.*, 66–72. Berlin: Springer.

Dahlke, S., and P. Maaß. 1997. “Interpolating refinable functions and wavelets for general scaling matrices.” *Numer. Funct. Anal. Optim.* 18 (5&6): 521–539.

Dahlke, S., P. Maaß, and G. Teschke. 2003. “Interpolating scaling functions with duals.” *Journal of Computational Analysis and Applications* 5 (3): 361–273.

Dahmen, W. 1994. “Some remarks on multiscale transformations, stability, and biorthogonality.” In *Wavelets, images, and surface fitting (Chamonix-Mont-Blanc, 1993)*, 157–188. Wellesley, MA: A K Peters.

———. 1995. “Multiscale analysis, approximation, and interpolation spaces.” In *Approximation theory VIII, Vol. 2 (College Station, TX, 1995)*, Volume 6 of *Ser. Approx. Decompos.*, 47–88. River Edge, NJ: World Sci. Publishing.

———. 1996. “Stability of multiscale transformations.” *J. Fourier Anal. and Appl.*, vol. 2.

———. 1997. “Wavelet and multiscale methods for operator equations.” *Acta Numerica* 6:55–228.

———. 2001. “Wavelet Methods for PDEs - Some Recent Developments.” *J. Comp. Appl. Math.* 128:123–185.

Dahmen, W., and A. Kunoth. 1992. "Multilevel preconditioning." *Numer. Math.* 63 (3): 315–344.

Dahmen, W., and C.A. Micchelli. 1983. "Translates of multivariate splines." *Linear Algebra and its Applications* 52:217–234.

Daubechies, I. 1992. *Ten lectures on wavelets.* Volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

de Boor, C. 1987. "The polynomials in the linear span of integer translates of a compactly supported function." *Constr. Approx.* 3 (2): 199–208.

———. 1990. "Quasiinterpolants and approximation power of multivariate splines." In *Computation of curves and surfaces (Puerto de la Cruz, 1989)*, Volume 307 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, 313–345. Dordrecht: Kluwer Acad. Publ.

de Boor, C., R. A. DeVore, and A. Ron. 1993. "On the construction of multivariate (pre)wavelets." *Constr. Approx.* 9 (2-3): 123–166.

———. 1994. "The structure of finitely generated shift-invariant spaces in $L_2(\mathbf{R}^d)$." *J. Funct. Anal.* 119 (1): 37–78.

de Boor, C., and G. Fix. 1973. "Spline approximation by quasi-interpolants." *J. Approx. Theory* 8:19–45.

Deny, J., and J.L. Lions. 1954. "Les espaces de type de Beppo Levi." *Ann. Inst. Fourier* 5:305–370.

Derado, J. 1999. "Multivariate refinable interpolating functions." *Appl. Comput. Harmon. Anal.* 7 (2): 165–183.

———. 2001. "Nonseparable, compactly supported interpolating refinable functions with arbitrary smoothness." *Appl. Comput. Harmon. Anal.* 10 (2): 113–138.

DeVore, R.A. 1998. "Nonlinear approximation." *Acta Numerica* 7:51–150.

DeVore, R.A., B. Jawerth, and V. Popov. 1992. "Compression of wavelet decompositions." *Am. J. Math.* 114 (4): 737–785.

DeVore, R.A., and G.G. Lorentz. 1993. *Constructive approximation.* Grundlehren der Mathematischen Wissenschaften. 303. Berlin: Springer- Verlag.

DeVore, R.A., and B.J. Lucier. 1992. "Wavelets." *Acta Numerica* 1:1–56.

DeVore, R.A., and V.A. Popov. 1988. "Interpolation of Besov spaces." *Trans. Am. Math. Soc.* 305 (1): 397–414.

Ditzian, Z., and V. Totik. 1987. *Moduli of smoothness.* Volume 9. New York: Springer Series in Computational Mathematics.

Dubinin, V.V. 1997. "Greedy Algorithms and applications." Ph.D. diss., University of South Carolina.

Feilner, M., D. Van De Ville, and M. Unser. "An Orthogonal Family of Quincunx Wavelets with Continuously-Adjustable Order." in press.

Garrigós, G., R. Hochmuth, and A. Tabacco. 2004. "Wavelet characterizations for anisotropic Besov spaces with $0 < p < 1$." *Proc. Edinb. Math. Soc. (2)* 47 (3): 573–595.

Garrigós, G., and A. Tabacco. 2002. "Wavelet decompositions of anisotropic Besov spaces." *Math. Nachr.* 239/240:80–102.

Goodman, T. N. T., R.-Q. Jia, and D.-X. Zhou. 2000. "Local linear independence of refinable vectors of functions." *Proc. Roy. Soc. Edinburgh Sect. A* 130 (4): 813–826.

Gröchenig, K., and W. R. Madych. 1992. "Multiresolution analysis, Haar bases, and self-similar tilings of $\mathbf{R}^n$." *IEEE Trans. Inform. Theory* 38 (2, part 2): 556–568.

Han, B. 2002. "Symmetry property and construction of wavelets with a general dilation matrix." *Linear Algebra Appl.* 353 (1-3): 207–225.

———. 2003. "Compactly supported tight wavelet frames and orthonormal wavelets of exponential decay with a general dilation matrix." *J. Comput. Appl. Math.* 155 (1): 43–67.

Han, B., and R.-Q. Jia. 1998. "Multivariate Refinement Equations and Subdivision Schemes." *SIAM J. Math. Anal.* 29 (5): 1177–1199.

———. 2002. "Quincunx fundamental refinable functions and quincunx biorthogonal wavelets." *Math. Comput.* 71 (237): 165–196.

He, W., and M.J. Lai. 2000. "Examples of Bivariate Nonseparable Compactly Supported Orthonormal Continuous Wavelets." *IEEE Transactions on Image Processing* 9:949–953.

Hochmuth, R. 2002a. "n-term approximation in anisotropic function spaces." *Math. Nachr.* 244:131–149.

———. 2002b. "Wavelet characterizations for anisotropic Besov spaces." *Appl. Comput. Harmon. Anal.* 12 (2): 179–208.

Holschneider, M. 1995. *Wavelets: An Analysis Tool.* Oxford University Press.

Jia, R.-Q. 1995. "The Toeplitz theorem and its applications to approximation theory and linear PDEs." *Trans. Amer. Math. Soc.* 347 (7): 2585–2594.

———. 1996. "The subdivision and transition operator assiciated with a refinement equation." *Advanced Topics in Multivariate Approximation*, pp. 139–154.

———. 1998. "Approximation Properties of Multivariate Wavelets." *Mathematics of Computation* 67 (222): 667–665 (April).

———. 1999. "Characterization of smoothness of multivariate refinable functions in sobolev spaces." *Trans. Am. Math. Soc.* 351 (10): 4089–4112.

Jia, R.-Q., and J. Lei. 1993. "Approximation by multiinteger translates of functions having global support." *J. Approximation Theory* 72 (1): 2–23.

Jia, R.-Q., and C. A. Micchelli. 1992. "Using the refinement equation for the construction of pre-wavelets. V. Extensibility of trigonometric polynomials." *Computing* 48 (1): 61–72.

Jia, R.-Q., and C.A. Micchelli. 1991. "Using the refinement equations for the construction of prewavelets. II: Powers of two." In *Curves and surfaces, Pap. Int. Conf., Chamonix-Mont-Blanc/Fr. 1990* , 209–246. Academic Press, New York.

Jia, R.-Q., and S. Zhang. 1999. "Spectral Properties of the Transition Operator Associated to a Multivariate Refinement Equation." *Linear Algebra Appl.* 292 (1-3): 155–178.

Johnen, H., and K. Scherer. 1977. "On the equivalence of the K-functional and moduli of continuity and some applications." *Constr. Theory Funct. several Variables, Proc. Conf. Oberwolfach 1976, Lect. Notes Math. 571.* Springer Verlag, 119–140.

Kovačević, J., and M. Vetterli. 1992. "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathbb{R}^n$." *IEEE Trans. Inform. Theory* 38 (2, part 2): 533–555.

Kunoth, A. 1994. "Multilevel preconditioning." Ph.D. diss., RWTH Aachen.

Latour, V., J. Müller, and W. Nickel. 1998. "Stationary subdivision for general scaling matrices." *Math. Z.* 227 (4): 645–661.

Lemari-Rieusset, P.-G., and G. Malgouyres. 1991. "Support des fonctions de base dans une analyse multi-rsolution." *CRAS Paris Série I* 213:377–380.

Lin, E.B., and Y. Ling. 2003. "Image compression and denoising via nonseparable wavelet approximation." *J: Comp. Appl. Math.* 155:131–152.

Mallat, S. 1998. *A wavelet tour of signal processing.* San Diego, CA: Academic Press. xxiv.

Mendivil, F., and D. Piché. 1999. "Two algorithms for non-separable wavelet transforms and applications to image compression." In *Fractals: Theory and applications in engineering.*, edited by M. Dekking et al., 325–345. London: Springer.

Meyer, Y. 1992. *Wavelets and operators. Translated by D. H. Salinger.* Cambridge Studies in Advanced Mathematics. 37. Cambridge: Cambridge University Press. xv.

Mojsilovic, A., S. Markovic, and M. Popovic. 1997. "Texture Analysis and Classification with the Nonseparable Wavelet Transform." *Proc. 1997 IEEE Int. Conf. on Image Processing, Santa Barbara*, Volume 3. 182–185.

Mojsilovic, A., and M. Popovic. 1998. "Characterization of visually similar diffuse diseases from B-scan liver images using the nonseparable wavelet transform." *IEEE Trans. on Medical Imaging* 17, no. 4.

Oswald, P., and E.A. Storozhenko. 1978. "Theorem of Jackson in the spaces $L_p(\mathbb{R}^k), 0 < p < 1$." *Siberian Math. J.* 19 (4): 888–901.

Roman, S. 1980. "The Formula of Faa di Bruno." *Amer. Math. Monthly* 87 (10): 805–809.

Runst, T., and W. Sickel. 1996. *Sobolev spaces of fractional order, Nemytskij operators and nonlinear partial differential equations.* de Gruyter Series in Nonlinear Analysis and Applications. 3. Berlin: de Gruyter. x.

Stein, E.M., and G. Weiss. 1971. *Introduction to Fourier analysis on Euclidean spaces.* Princeton Mathematical Series. Princeton, N. J.: Princeton University Press. X.

Strang, G., and T. Nguyen. 1997. *Wavelets and Filter banks.* Wellesley-Cambridge Press.

Sun, Q. 2001. "Linear independence of the integer translates of compactly supported distributions and refinable vectors." *Appl. Math. J. Chinese Univ. Ser. B* 16 (4): 381–396.

Temlyakov, V.N. 1998a. "The best m-term approximation and greedy algorithms." *Adv. Comput. Math.* 8 (3): 249–265.

———. 1998b. "Nonlinear m-term approximation with regard to the multivariate Haar system." *East J. Approx.* 4 (1): 87–106.

Triebel, H. 1983. *Theory of Function Spaces.* Volume 78. Basel-Boston-Stuttgart: Monographs in Mathematics, Birkhäuser Verlag.

———. 1992. *Theory of function spaces II.* Monographs in Mathematics. 84. Basel etc.: Birkhuser Verlag. viii .

Van der Auwera, G., A. Munteanu, and J. Cornelis. 2000. "Evaluation of a Quincunx Wavelet Filter Design Approach for Quadtree-Based Embedded Image Coding." *IEEE International Conference on Image Processing, ICIP2000, Vancouver, Canada,*, Volume 3. 190–194.

van de Wouwer, G. 1998. Non-separable Wavelets for Rotation-Invariant Texture Classification and Segmentation (part of Phd thesis).

Vetterli, M., and J. Kovačević. 1995. *Wavelets and Subband Coding.* New Jersey: Prentice Hall PTR.

Wojtaszczyk, P. 1997. *A mathematical introduction to wavelets.* Cambridge: Cambridge University Press.